

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:25:41 ; Search time 20.636 Seconds
(without alignments)
2556.865 Million cell updates/sec

Title: US-08-325-278B-1
Perfect score: 1565
Sequence: 1 AVENKEETPTDSEEV.....GGYTINIRFAGKVKDEKPEE 305

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL19.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	1565	100.0	455	2 Q53291	Q53291 streptococc
2	1561	99.7	719	2 Q51912	Q51912 peptostrept
3	1226.5	78.4	992	2 Q51918	Q51918 peptostrept
4	167	10.7	664	2 Q53975	Q53975 streptococc
5	166	10.6	479	10 Q9LF88	Q9LF88 arabidopsis
6	166	10.6	669	2 Q9SEM8	Q9SEM8 streptococc
7	151	9.6	448	10 Q9SKP0	Q9SKP0 arabidopsis
8	149	9.5	448	10 Q96246	Q96246 arabidopsis
9	148.5	9.5	2276	2 Q93TY6	Q93TY6 streptococc
10	141.5	9.0	1110	13 Q91255	Q91255 streptococc
11	141	9.0	1849	2 Q9SAK2	Q9SAK2 lactobacill
12	137.5	8.8	992	16 Q9WZA6	Q9WZA6 thermotoga
13	137.5	8.8	5458	5 Q90459	Q90459 plasmodium
14	136	8.7	506	2 Q53837	Q53837 salmonella
15	134.5	8.6	3198	5 Q9U8G8	Q9U8G8 manduca sex
16	132.5	8.5	737	16 Q9JYK4	Q9JYK4 neisseria m

17	132.5	8.5	17352	5 Q95YM2	Q95YM2 procamburus
18	131.5	8.4	891	2 Q04111	Q04111 enterococcu
19	131.5	8.4	1166	2 Q86489	Q86489 staphylococ
20	130.5	8.3	564	2 Q89188	Q89188 streptococ
21	130.5	8.3	581	2 Q33742	Q33742 streptococ
22	130.5	8.3	663	2 Q30874	Q30874 streptococ
23	130.5	8.3	701	2 Q9KK48	Q9KK48 streptococ
24	130.5	8.3	1141	16 Q9W46	Q9W46 staphylococ
25	130.5	8.3	1141	16 Q932F7	Q932F7 staphylococ
26	129.5	8.3	1733	2 Q9K114	Q9K114 staphylococ
27	129	8.2	1171	2 Q9KW6	Q9KW6 staphylococ
28	128.5	8.2	938	16 Q927R4	Q927R4 listeria in
29	128.5	8.2	1119	13 P87344	P87344 theagra ch
30	127.5	8.1	701	2 Q9R0T5	Q9R0T5 streptococ
31	127.5	8.1	717	2 Q9EY85	Q9EY85 clostridium
32	127.5	8.1	728	2 Q9Z4N2	Q9Z4N2 neisseria g
33	127.5	8.1	843	2 Q47802	Q47802 enterococu
34	127	8.1	913	13 Q13099	Q13099 xenopus lae
35	126.5	8.1	1035	5 Q21380	Q21380 caenorhabdi
36	126.5	8.1	2045	16 Q9A0K5	Q9A0K5 streptococ
37	126	8.1	955	2 Q45574	Q45574 bacillus sp
38	125.5	8.0	673	10 Q9W9S5	Q9W9S5 arabidopsis
39	125.5	8.0	1510	5 Q25920	Q25920 plasmodium
40	125	8.0	463	10 Q39871	Q39871 glycine max
41	125	8.0	2083	5 Q9N435	Q9N435 caenorhabdi
42	124.5	8.0	1302	2 Q49547	Q49547 mycoplasma
43	124	7.9	495	16 Q9W6G8	Q9W6G8 staphylococ
44	124	7.9	506	2 Q53834	Q53834 salmonella
45	124	7.9	1489	10 Q96449	Q96449 phytophthor

ALIGNMENTS

RESULT 1

Q53291 ID Q53291 PRELIMINARY; PRT; 455 AA.
AC Q53291;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 01, Last annotation update)
DE PROTEIN LG (FRAGMENT).
OS Streptococcus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93094283; PubMed=1460053;
RA Kihlberg B.M., Sjöbring U., Kaster W., Björck L.;
RT "Protein LG: a hybrid molecule with unique immunoglobulin binding
RT properties";
RL J. Biol. Chem. 267:25583-25588(1992). Dec 1992
DR EMBL; SS0809; AAA03280.1; -
DR HSSP; P06654; LPGA.
DR InterPro; IPR003147; B1.
DR InterPro; IPR000724; Igg_bind_B.
DR Pfam; PF02246; B1; 4.
DR Pfam; PF01378; Igg_binding_B; 2.
FT NON_TER 455 455
SQ SEQUENCE 455 AA; 49926 MW; 381FC235BBC8307B CRC64;

Query Match 100.0%; Score 1565; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 9, le-79;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AVENKEETPTDSEEV	VTKANLIFANGSTQTEFEKGTPEKATSEAYAYADTLKDN	60
Db	22	AVENKEETPTDSEEV	VTKANLIFANGSTQTEFEKGTPEKATSEAYAYADTLKDN	81
QY	61	GEYTVDAVGKVTNLKFKAGKTPPEKPEEVT	IRKANLIYADGKTQTAEFKGTFEEATAE	120
Db	82	GEYTVDAVGKVTNLKFKAGKTPPEKPEEVT	IRKANLIYADGKTQTAEFKGTFEEATAE	141

THIS PAGE BLANK (uspro)

QY 121 AYRYADALKKDNQGYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 180
|||||
Db 142 AYRYADALKKDNQGYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 201
|||||
QY 181 EKGTFEEATAEAYRYADLLAKENGYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKA 240
|||||
Db 202 EKGTFEEATAEAYRYADLLAKENGYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKA 261
|||||
QY 241 NLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTNIRFAGKKYD 300
|||||
Db 262 NLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTNIRFAGKKYD 321
|||||
QY 301 ERPEE 305
|||||
Db 322 ERPEE 326
|||||

RESULT 2
Q51912
ID Q51912 PRELIMINARY; PRT; 719 AA.
AC Q51912;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE PROTEIN L PRECURSOR.
OS Peptostreptococcus magnus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Finegoldia.
OX NCBI_TaxID=1260;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=312;
RX MEDLINE=90215984; PubMed=2108927;
RA Kastern W., Holst E., Nielsen E., Sjoerding U., Bjorck L.;
RT "Protein L, a bacterial immunoglobulin-binding protein and possible
RT virulence determinant.";
RL Infect. Immun. 58:1217-1223(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=312;
RX MEDLINE=92316971; PubMed=1618782;
RA Bjoerck L., Sjoerding U., Kastern W.;
RT "Structure of peptostreptococcal protein L and identification of
RT repeated immunoglobulin light chain-binding domain.";
RL J. Biol. Chem. 267:12820-12825(1992).
DR EMBL; M86697; AAA25612.1;
DR InterPro; IPR003147; Bl.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF02246; Bl; 5.
DR Pfam; PF00746; Gram_pos_anchor; 1.
KW Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 719 PROTEIN L.
SQ SEQUENCE 719 AA; 78983 MW; 963A8D76D5E34DD2 CRC64;

Query Match 99.7%; Score 1561; DB 2; Length 719;
Best Local Similarity 100.0%; Pred. No. 2.5e-78;
Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VENKEETPETDSEEEVTIKANLIANGSTQTAFFKGTFFKATSEAYAYADTLKKDNG 61
|||||
Db 95 VENKEETPETDSEEEVTIKANLIANGSTQTAFFKGTFFKATSEAYAYADTLKKDNG 154
|||||
QY 62 EYTVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAFFKGTFFKATSEAY 121
|||||
Db 155 EYTVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAFFKGTFFKATSEAY 214
|||||
QY 122 YRYADALKKDNQGYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 181
|||||
Db 155 YRYADALKKDNQGYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 274
|||||
QY 182 FKGTFEEATAEAYRYADLLAKENGYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKAN 241
|||||

Db 275 FKGTFEEATAEAYRYADLLAKENGYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKAN 334
|||||
QY 242 LIYADGKTQTAFFKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTNIRFAGKKYD 301
|||||
Db 335 LIYADGKTQTAFFKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTNIRFAGKKYD 394
|||||
QY 302 KPEE 305
|||||
Db 395 KPEE 398
|||||
RESULT 3
Q51918
ID Q51918 PRELIMINARY; PRT; 992 AA.
AC Q51918;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE PROTEIN L PRECURSOR.
OS Peptostreptococcus magnus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Finegoldia.
OX NCBI_TaxID=1260;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3316;
RX MEDLINE=95078460; PubMed=7987012;
RA Murphy J.P., Trowern A.R., Dugleby C.J.;
RT "Nucleotide sequence of the gene for peptostreptococcal protein L.";
RL DNA Seq. 4:259-265(1994).
DR EMBL; L04466; AAA67503.1;
DR HSSP; Q51911; LGAB.
DR InterPro; IPR003147; Bl.
DR InterPro; IPR002988; GA.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF02246; Bl; 4.
DR Pfam; PF01468; GA; 4.
DR Pfam; PF00746; Gram_pos_anchor; 1.
KW Signal.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 992 PROTEIN L.
SQ SEQUENCE 992 AA; 108700 MW; 9CFF5771578A5DCE CRC64;

Query Match 78.4%; Score 1226.5; DB 2; Length 992;
Best Local Similarity 79.9%; Pred. No. 8.5e-60;
Matches 246; Conservative 22; Mismatches 31; Indels 9; Gaps 4;

QY 2 VENKEETPETDSEEEVTIKANLIANGSTQTAFFKGTFFKATSEAYAYADTLKKDNG 61
|||||
Db 241 MERKLESEKTEP--PEEVTIKANLIADGSGTQNAFFKGTFAKAVSDAYADALKDNG 298
|||||
QY 62 EYTVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAFFKGTFFKATSEAY 121
|||||
Db 299 EYTVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAFFKGTFFKATSEAY 358
|||||
QY 122 YRYADALKKDNQGYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 179
|||||
Db 359 YAYADLLAKENGYTADLEDGGYTNIRFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 418
|||||
QY 180 AEFKGTFFKATSEAYRYADLLAKENGYTVDVADKGYTLNIFAGKEKTPPEPKEEVTI 237
|||||
Db 419 AEFKGTFFKATSEAYRYADLLAKENGYTADLEDGGYTNIRFAGKEKTPPEPKEEVTI 478
|||||
QY 238 IKANLIYADGKTQTAFFKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTNIRFAGK 297
|||||
Db 479 IKVNLIFADGKTQTAFFKGTFFKATSEAYRYADLLAKENGYTADLEDGGYTNIRFAGK 538
|||||
QY 298 KVDEKPEE 305
|||||
Db 539 ---EQPGE 543
|||||

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:25:41 ; Search time 29.364 seconds
(without alignments)
2556.865 Million cell updates/sec

Title: US-08-325-278B-3
Perfect score: 2235
Sequence: 1 AVENKEETPETDSEEV.....GVDGWTYDDATKTFVTVM 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTEMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	2235	100.0	455	2	Q53291	Q53291 streptococc
2	1587.5	71.0	719	2	Q51912	Q51912 peptostrept
3	1263.5	56.5	992	2	Q51918	Q51918 peptostrept
4	706.5	31.6	664	2	Q53975	Q53975 streptococc
5	697.5	31.2	669	2	Q938M8	Q938M8 streptococc
6	670	30.0	185	2	Q54181	Q54181 streptococc
7	481	21.5	429	2	Q56212	Q56212 streptococc
8	404	18.1	413	2	Q53974	Q53974 streptococc
9	294	13.2	60	2	Q53337	Q53337 streptococc
10	182.5	8.2	3198	5	Q9UG88	Q9UG88 manduca sex
11	182	8.1	479	10	Q9LF88	Q9LF88 arabidopsis
12	177.5	7.9	2402	2	Q9AER7	Q9AER7 staphylococ
13	175	7.8	448	10	Q9SKP0	Q9SKP0 arabidopsis
14	170.5	7.6	1849	2	Q9SAK2	Q9SAK2 lactobacilli
15	170	7.6	2276	2	Q93TY6	Q93TY6 staphylococ
16	168	7.5	448	10	Q96246	Q96246 arabidopsis

17	162.5	7.3	2353	2	P71401	P71401 haemophilus
18	162	7.2	913	13	O13099	O13099 xenopus lae
19	161	7.2	2692	4	Q9Y503	Q9Y503 homo sapien
20	160.5	7.2	5458	5	Q9U459	Q9U459 plasmodium
21	160	7.2	2691	4	Q9NS12	Q9NS12 homo sapien
22	160	7.2	2705	4	Q9S303	Q9S303 homo sapien
23	160	7.2	2725	4	Q9NYE5	Q9NYE5 homo sapien
24	159	7.1	1856	16	Q97T80	Q97T80 streptococc
25	158	7.1	2045	16	Q9AOK5	Q9AOK5 streptococc
26	157.5	7.0	1166	2	O86489	O86489 staphylococ
27	157	7.0	798	13	Q90307	Q90307 carassius a
28	156.5	7.0	564	2	O69188	O69188 streptococc
29	156.5	7.0	581	2	O33742	O33742 streptococc
30	156.5	7.0	663	2	O30874	O30874 streptococc
31	156.5	7.0	701	2	Q9KK48	Q9KK48 streptococc
32	156.5	7.0	969	5	Q9NDI9	Q9NDI9 plasmodium
33	156	7.0	1368	2	Q93M90	Q93M90 clostridium
34	154.5	6.9	1110	13	Q91255	Q91255 petromyzon
35	154.5	6.9	1822	2	Q07290	Q07290 streptococc
36	154	6.9	737	16	Q9JYK4	Q9JYK4 neisseria m
37	154	6.9	845	5	Q9Y1P8	Q9Y1P8 plasmodium
38	154	6.9	1698	2	Q9LC00	Q9LC00 staphylococ
39	153.5	6.9	492	2	Q56193	Q56193 staphylococ
40	153.5	6.9	573	2	Q56192	Q56192 staphylococ
41	153.5	6.9	701	2	Q9RQT5	Q9RQT5 streptococc
42	153.5	6.9	1873	2	Q9Z4N7	Q9Z4N7 enterococcu
43	153	6.8	897	13	O13098	O13098 xenopus lae
44	152	6.8	3147	5	Q17464	Q17464 caenorhabdi
45	151.5	6.8	717	2	Q9EY85	Q9EY85 clostridium

ALIGNMENTS

RESULT 1

ID	Q53291	PRELIMINARY;	PRT;	455 AA.
AC	Q53291;			
DT	01-NOV-1996 (TRENBLrel. 01, Created)			
DT	01-NOV-1996 (TRENBLrel. 01, Last sequence update)			
DT	01-DEC-2001 (TRENBLrel. 19, Last annotation update)			
DE	PROTEIN LG (FRAGMENT).			
OS	Streptococcus sp.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
OC	Streptococcus			
OX	NCBI_TaxID=1306;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93094283; PubMed=1460053;			
RA	Kihlberg B.M., Sjöbring U., Kastern W., Björck L.;			
RT	"Protein LG: a hybrid molecule with unique immunoglobulin binding			
RT	properties."			
RL	J. Biol. Chem. 267:25583-25588(1992).			
DR	EMBL; S50809; AAA03280.1; -.			
DR	HSNP; P06654; 1PGX.			
DR	InterPro; IPR003147; BL.			
DR	InterPro; IPR000724; IgG_bind_B.			
DR	Pfam; PF02246; BL; 4.			
DR	Pfam; PF01378; IgG_binding_B; 2.			
FT	NON_TER 455 455			
SQ	SEQUENCE 455 AA; 49926 MW; 381FC235BBC8307B CRC64;			

Query Match	100.0%;	Score	2235;	DB	2;	Length	455;
Best Local Similarity	100.0%;	Pred. No.	1e-100;				
Matches	434;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	AVENKEETPETDSEEVTKANLIFANGSTOTAEFKGTFEKATSEAYAYADTLKKDN	60				

```
QY 121 AYRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 180
|||||
Db 142 AYRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 201
|||||
QY 181 EFKGTPEEATAEAYRYADLLAKENGYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKA 240
|||||
Db 202 EFKGTPEEATAEAYRYADLLAKENGYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKA 261
|||||
QY 241 NLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTNIRFAGKKVD 300
|||||
Db 262 NLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTNIRFAGKKVD 321
|||||
QY 301 EKPEEPMDTYKLLINGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGWTYDDATKTFT 360
|||||
Db 322 EKPEEPMDTYKLLINGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGWTYDDATKTFT 381
|||||
QY 361 VTEKPEVIDASELTPAVTTYKLVINGKTLKGETTTRAVDAETAFAKQYANDNGVDGVW 420
|||||
Db 382 VTEKPEVIDASELTPAVTTYKLVINGKTLKGETTTRAVDAETAFAKQYANDNGVDGVW 441
|||||
QY 421 TYDDATKTFTVTEM 434
|||||
Db 442 TYDDATKTFTVTEM 455
|||||
```

```
RESULT 2
Q51912
ID Q51912 PRELIMINARY; PRT; 719 AA.
AC Q51912;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PROTEIN L PRECURSOR.
OS Peptostreptococcus magnus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Finegoldia.
OX NCBI_TaxID=1260;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=312;
RX MEDLINE=90215984; PubMed=2108927;
RA Kastern W., Holst E., Nielsen E., Sjoerck U., Bjorck L.;
RA "Protein L, a bacterial immunoglobulin-binding protein and possible
RT virulence determinant.";
RL Infect. Immun. 58:1217-1222(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=312;
RX MEDLINE=92316971; PubMed=1618782;
RA Bjoerck L., Sjoerck U., Kastern W.;
RA "Structure of peptostreptococcal protein L and identification of
RT repeated immunoglobulin light chain-binding domain.";
RL J. Biol. Chem. 267:12820-12825(1992).
DR EMBL; M86697; AAA25612.1; -.
DR InterPro; IPR003147; B1.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF02246; B1; 5.
DR Pfam; PF00746; Gram_pos_anchor; 1.
KW Signal.
FT SIGNAL.
FT CHAIN.
FT CHAIN 1 18 POTENTIAL.
FT CHAIN 19 719 PROTEIN L.
SQ SEQUENCE 719 AA; 78983 MW; 963A8D76D5E34DD2 CRC64;
```

Query Match 71.0%; Score 1587.5; DB 2; Length 719;
Best Local Similarity 76.3%; Pred. No. 3e-69;
Matches 334; Conservative 13; Mismatches 56; Indels 35; Gaps 6;

```
QY 2 VENKEETPETPDSEEVTIKANLIFANGSTQTAEFKGTFAEAYADTLKKDNG 61
|||||
Db 95 VENKEETPETPDSEEVTIKANLIFANGSTQTAEFKGTFAEAYADTLKKDNG 154
|||||
QY 62 EYTDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTAEFKGTFAEATAEA 121
```

```
Db 155 EYTDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTAEFKGTFAEATAEA 214
|||||
QY 122 YRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 181
|||||
Db 215 YRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 274
|||||
QY 182 FKGTPPEATAEAYRYADLLAKENGYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKAN 241
|||||
Db 275 FKGTPPEATAEAYRYADLLAKENGYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKAN 334
|||||
QY 242 LIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTNIRFAGKKVDE 301
|||||
Db 335 LIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTNIRFAGKKVDE 394
|||||
QY 302 KPEEPM-D-TYK--LIINGKTLKGETTTEAVDAATAEKVFKQYANDNGVD-GEWYDDATK 357
|||||
Db 395 KPEEKEQVTIKENIYFEDGTQATFKGTFAEATAEA--RYADLLSKEHGYTADLEDG 452
|||||
QY 358 TPTVT-----EKPEVIDASELTPAVTTYKLVINGKTLKGETTTRAVDAETAFAK 404
|||||
Db 453 GYTNIRFAGKEPEETPEKPEVQD-----GYASYEEAAAKALKND 496
|||||
QY 405 KAFKQYANDNGVDGVWY 422
|||||
Db 497 DVNKSYYTIRGADGRYY 514
```

```
RESULT 3
Q51918
ID Q51918 PRELIMINARY; PRT; 992 AA.
AC Q51918;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PROTEIN L PRECURSOR.
OS Peptostreptococcus magnus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Finegoldia.
OX NCBI_TaxID=1260;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3316;
RX MEDLINE=95078460; PubMed=7987012;
RA Murphy J.P., Trovern A.R., Duggleby C.J.;
RA "Nucleotide sequence of the gene for peptostreptococcal protein L.";
RL DNA Seq. 4:259-265(1994).
DR EMBL; L04466; AAA67503.1; -.
DR HSSP; Q51911; 1GAB.
DR InterPro; IPR003147; B1.
DR InterPro; IPR002988; GA.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF02246; B1; 4.
DR Pfam; PF01468; GA; 4.
DR Pfam; PF00746; Gram_pos_anchor; 1.
KW Signal.
FT SIGNAL.
FT CHAIN.
FT CHAIN 1 24 POTENTIAL.
FT CHAIN 25 992 PROTEIN L.
SQ SEQUENCE 992 AA; 108700 MW; 9CFF5771578A5DCE CRC64;
```

Query Match 56.5%; Score 1263.5; DB 2; Length 992;
Best Local Similarity 58.8%; Pred. No. 1.8e-53;
Matches 281; Conservative 36; Mismatches 90; Indels 71; Gaps 12;

```
QY 2 VENKEETPETPDSEEVTIKANLIFANGSTQTAEFKGTFAEAYADTLKKDNG 61
|||||
Db 241 MERKLEKETPE--PEEYTIKANLIFADGQTQNAEFKGTFAKAVSDAYADALKKNG 298
|||||
QY 62 EYTDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTAEFKGTFAEATAEA 121
|||||
Db 299 EYTDVADKGYTLNKFAGKEKTPPEPKKEEVTIKVNLIFADGKTQTAEFKGTFAEATAKA 358
|||||
QY 122 YRYADALKKDNGETYVDVADKGYTLNKFAGKE--KTPEEPEEVTIKANLIYADGKTQ 179
|||||
```

FUNCTION.

RP MEDLINE=99081771; PubMed=9864244;

RA Vasi J., Svensson J., Frick I.-M., Mueller H.-P.;

RT "Five homologous repeats of the protein G-related protein MIG

RT cooperate in binding to goat immunoglobulin G.,"

RL Infect. Immun. 67:413-416(1999).

CC -1- FUNCTION: BINDS TO THE CONSTANT FC REGION OF IGG WITH HIGH

CC AFFINITY. N-TERMINUS MEDIATES BINDING TO PLASMA PROTEINASE

CC INHIBITOR ALPHA 2-MACROGLOBULIN AFTER COMPLEX FORMATION WITH

CC PROTEASES.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL (BY

CC SIMILARITY).

CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS IN

CC THE REGION OF THE MEMBRANE ANCHOR. C-TERMINAL IS SIMILAR TO OTHER

CC STREPTOCOCCAL TYPE-III FC RECEPTORS.

DR EMBL: Z29666; CAA82764.1; "

DR HSPS: P06654; 1PGX

DR InterPro: IPR001899; Gram_pos_anchor.

DR InterPro: IPR000724; IgG_bind_B.

DR InterPro: IPR000531; TonB_boxC.

DR Pfam: PF00746; Gram_pos_anchor; 1.

DR Pfam: PF01378; IgG_binding_B; 5.

DR PRINTS: PR00015; GPOSANCHOR.

DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.

DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.

DR IgG-binding protein; Repeat; Transmembrane; Cell wall; Signal.

KW

```

RESULT 5
Q93EM8
ID Q93EM8 PRELIMINARY; PRT; 669 AA.
AC Q93EM8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MIG PRECURSOR.
OS Streptococcus dysgalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1334;
[1]
RN SEQUENCE FROM N.A.

```



```
Db 141 DAASVDQVNAAIN-----DAHTAIADITGA---ALLFAKEAAINELKOYGISDYVVTLINK 193
QY 287 GYTI-----NIRFAGKKVDEKPEP-MDTYKLIINGKTLKGETTTEAVDAATAE 334
Db 194 AKTVEGVNALKAEILSALPSSEVIDAELTPALTSKVLKIGATFGSETATKAVDRAVAE 253
QY 335 KVEFOYANDGVGENTYDDATKTFVTPEK--EVIDASLTPAVTYKLVINGKTLKAE 392
Db 254 QTPRDYANKNGVDGVWAYDAATKTFVTPEQVAETIEAAELTPALTYRLVIGKVTFSGE 313
QY 393 TTTKAVDAETAFAKFOYANDGVGWYDDATKTFVTPE 433
Db 314 TATKAVDAATAEAQFROYANDGVGTGEWAYDAATKTFVTPE 354

RESULT 8
Q53974
ID Q53974 PRELIMINARY; PRT; 413 AA.
AC Q53974;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DE MAG PROTEIN PRECURSOR.
GN MAG.
OS Streptococcus dysgalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1334;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94259307; PubMed=7515368;
RA Jonsson H., Frykberg L., Rantamaeki L., Guss B.;
RT "MAG, a novel plasma protein receptor from Streptococcus
  dysgalactiae.";
RL Gene 143:85-89(1994).
DR EMBL; L27798; AAA26921.1; -.
DR HSSP; P06654; IPGX.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000724; IgG_bind_B.
DR Pfam; PF01468; GA; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF01378; IgG_binding_B; 1.
DR PRINTS; PR00015; GPOANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Signal; Transmembrane.
FT SIGNAL 1 34 POTENTIAL.
FT CHAIN 35 413 POTENTIAL.
SQ SEQUENCE 413 AA; 44004 MW; F04DC71044F9E50F CRC64;

Query Match 18.1%; Score 404; DB 2; Length 413;
Best Local Similarity 32.3%; Pred. No. 1.9e-12;
Matches 141; Conservative 31; Mismatches 110; Indels 154; Gaps 14;

QY 1 AVENKEETPETDSEEVITKANLIFANGSTQAEFGKTFEKATSEAYAYADTLKKDN 60
Db 30 AVNAEESTVSPVTADAT-----TSKEALAIINKLSEDN 66
QY 61 GETVDVADKGYTLNKFAGKENTPEPKKEVITKANL-IYADGKQTQTAEFKGTFEATA 119
Db 67 LN-NLDI-----QEVLAkakADILASDSADT--IKALLAEVTA 101
QY 120 EAYRADALKNDGEYTVDVADKGYTLNKFAGKENTPEPKKEVITKANLIYADGKTQT 179
Db 102 EVTRL-----NEKMARDVADKAIADA-----AFSELKD 132
QY 180 AEFKGTPEEATAEAYRADL-LAKENGKTYVDVADKGYTLNKFAGKENTPEPKKEVTI 238
Db 133 AOLKAYEDLAKLAADTDLDAKAIINDYTTKVENAKTAEDVK-----KIFESQNEVT- 186
QY 239 KANLIYADGKTQTAEFKGTFAETAAYRADILLAKENGKYTADLDGGYTTINIRFAGKK 298
Db 187 -----RIKTEKALAAALAKAKADAIE-----ILKKYG-----IGDYVI----- 220
```

```
QY 299 VDEKPEPMDTYKLIINGKTLKGETTTEAVDAATAEAKFOYANDGVGENTYDDATKT 358
Db 221 -----KLINGKTAEGVT-----ALKD 237
QY 359 FVTEKPEVIDASELTPAVTYKLVINGKTLKGETTTTAKVDAETAFAKFOYANDGVG 418
Db 238 EILASKPAVIDAPELTPALTYKLVINGKTLKGETTTTAKVDAETAFAKFOYANDGVG 297
QY 419 VMTYDDATKTFVTTEM 434
Db 298 VMTYDDATKTFVTTEM 313

RESULT 9
Q53337
ID Q53337 PRELIMINARY; PRT; 60 AA.
AC Q53337;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DE PROTEIN G IGG FC BINDING DOMAIN.
OS Streptococcus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93277681; PubMed=8503988;
RA Cai S.Y., Wang Y.Y., Yao Z.J.;
RT "Structure analysis of streptococcal protein G Fc binding domain.";
RL Sci. China B 36:75-80(1993).
DR EMBL; S62801; AAB27024.1; -.
DR HSSP; P06654; IPGX.
DR InterPro; IPR000724; IgG_bind_B.
DR Pfam; PF01378; IgG_binding_B; 1.
SQ SEQUENCE 60 AA; 6655 MW; 924567E0D6B513DB CRC64;

Query Match 13.2%; Score 294; DB 2; Length 60;
Best Local Similarity 96.6%; Pred. No. 4.2e-08;
Matches 56; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 377 VTTYKLVINGKTLKGETTTTAKVDAETAFAKFOYANDGVGVTYDDATKTFVTTEM 434
Db 1 MTTYKLVINGKTLKGETTTTAKVDAETAFAKFOYANDGVGVTYDDATKTFVTTEM 58

RESULT 10
Q9U8G8
ID Q9U8G8 PRELIMINARY; PRT; 3198 AA.
AC Q9U8G8;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE LACUNIN PRECURSOR.
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Sphingioidea; Sphingidae; Sphinginae; Manduca.
OX NCBI_TaxID=7130;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99457716; PubMed=10528409;
RA Nardi J.B., Martos R., Walden K.K., Lampe D.J., Robertson H.M.;
RT "Expression of lacunin, a large multidomain extracellular matrix
  protein, accompanies morphogenesis of epithelial monolayers in Manduca
  sexta.";
RL Insect Biochem. Mol. Biol. 29:883-897(1999).
DR EMBL; AF078161; AAF04457.1; -.
DR HSSP; P12111; 2KNT.
DR InterPro; IPR004094; Antistasin.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003006; Ig_MHC.
```



```
DR InterPro: IPR003137; PA.
DR InterPro: IPR000209; Peptidase_s8.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 3.
DR PRINTS: PR00723; SUBTILISIN
DR PROSITE: PS00588; FLAGELIA_BB_ROD; UNKNOWN_1.
DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR PROSITE: PS00139; SUBTILASE_SER; 1.
SQ SEQUENCE 1849 AA; 199614 MW; 219F0D44B15A091F CRC64;

Query Match
Best Local Similarity 7.6%; Score 170.5; DB 2; Length 1849;
Matches 117; Conservative 44; Mismatches 165; Indels 125; Gaps 25;

QY 73 TLNKFAGKEKTEPEEKETIKANLIYADGKTQTAEFGKTFEEATAEA--YYADALKK 130
DB 1415 TLNL---DSENTYTNKKDKFTI-SGTISDDYKFYDLSINGNDVETSWSAVDYHSKEGIRK 1470
QY 131 DNGEYTVDVADKGYTLNLIK-----FAGKEKTEPEEKEE--VTIKANLIYA 173
DB 1471 -NPKHEVDLKKGNFTNVKVTIDIOGNSSOALVYVVEPAKTLAEPVVDKLLTKTANLQLL 1529
QY 174 DGKTQTAEFK-----GTFEATAEAYRYADLLAKENGKYTVDVADKGY-----TL 218
DB 1530 KATTDSEAKVWYSLDNGKTFNDVPADGFK-----VTENGTVQFKAVDK-YGNESKVKS 1583
QY 219 NIKFAGKEKTEPEEKETIKANL---IYADGKTQ----TAEFGKTFEATAEAYRYADL 271
DB 1584 EIKGLNKENQSPEDKELAKENLQAKVDAGEKKDLQKYTADSKDFNDALKKA----KDV 1640
QY 272 LAKENGKYTADLEDGGYTNINIRFAGKKVDEKPEEPMDTYKLIIL-NGKTLKGTTTEAVDA 330
DB 1641 LADKNAR-LADLOAKALD--RAEQALTEKPAEP--TIPLLOGNNVSNINTSSDNQV 1695
QY 331 ATAEEKVFKYANDN-----GVDGENTY-----DDATKTFT-----VTEKP--- 365
DB 1696 AAPVHAEKDKTKNDKNTTEGKDKTKVMFKSVLYTKDLKTRTAQAYSSLKLVTEKGK 1755
QY 366 -----EVIDASELTPA--VTYIKLVI-----NGKTL-----KGETTITKAV 398
DB 1756 VYTFKGHYFKVVDNRNAYVRNVGTGKATLKRNSFYQSGNGKSKRKLKKGTTITVYG 1815
QY 399 DATAEKAFKQYA-----NDNGVD 417
DB 1816 DQYKALKHYKKYAIRIGEGRYIKSVNVRVD 1846

RESULT 15
Q93TY6 PRELIMINARY; PRT; 2276 AA.
AC Q93TY6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BIOFILM-ASSOCIATED SURFACE PROTEIN.
GN BAP.
OS Staphylococcus aureus.
OC Bacteria, Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-V329;
RX MEDLINE=21189316; PubMed=11292810;
RA Cucarella C.; Solano C.; Valle J.; Amorena B.; Lasa I.;
RA Penades J.R.;
RT "Bap, a Staphylococcus aureus Surface Protein Involved in Biofilm
RT Formation.";
RL J. Bacteriol. 183:2888-2896(2001).
DR EMBL: AF288402; AAK38834.1; -.
SQ SEQUENCE 2276 AA; 238792 MW; 8623C60472CB0C0F CRC64;
```

```
Query Match
Best Local Similarity 7.6%; Score 170; DB 2; Length 2276;
Matches 110; Conservative 47; Mismatches 197; Indels 150; Gaps 25;

QY 8 TPETPETDSEEEVTTKANLIFANGSTQTAEFGKTFEATSEAYAYADTLKKNGEYTVDV 67
DB 1558 TSDATQVGTGAEPNSVTKLTFPDGTAT---GT-----AD-----DQGNYYTIDI 1598
QY 68 ADK-----GYTLNIRFAGKEKTEPEEK-----EETIKANLIYADGKTQT 107
DB 1599 PANGDLNGEELQVATDKDGNTPSESSANSNVDTTASDAPTVDVNTSDATQVTGAE 1658
QY 108 ABFFKGTFEATAEAYRYADALKKNGEYTVDVADK-----GYTLNIRFAGKEKTEPEEK- 161
DB 1659 T-VKLTFPDGTT-----ATGTADDQGNYYTIDIPSNVDLNGEELQVATDKDGNTPSE 1712
QY 162 -----BEVTTKANLIYADGKTQTAEFGKTFEATAEAYRYADLLAKENGKY 207
DB 1713 ANVTDTTAPDAPTVDVNTSDATQVTGAEPNST-VKLTFPDGTT-----ATGTADDQGN 1766
QY 208 TVDVADK-----GYTLNIRFAGKEKTEPEEK-----EETIKANLIYADG 247
DB 1767 TIDIPSNVDLNGEELQVATDKDGNTPSEPKLTNVDTTASDAPTVDVNTSDASQV 1826
QY 248 KTQTAEFGKTFEATAEAYRYADLLAKENGKYTABLE-----DGGYTNIRFAGKKVDEK 302
DB 1827 EPNST-VKLTFPDGTT-----ATGTADDQGNYYTIDIPSNVDLNGEELQVATDK- 1878
QY 303 PEEPMDTYKLIINGKTLKGTTTEAVDAATAEKVFKYQYANDNGVGEWYDDATKTFT 362
DB 1879 TSEPSA-----NVTDTTAPDAPTVDV-----TSDATQ---VT 1909
QY 363 EKPEVIDASELT-PAVTT-----YKLVI--NGKTLKGE---TTTKAVDAATAEKA 406
DB 1910 GGAEPNSTVKLTFPDGTTATGTADDQGNYYTIDIPSNVDLNGEELQVATDKDGNTPSE 1969
QY 407 FKQYANDNGVGEWYDDATKTFT 430
DB 1970 SANVDTTAPDAP-TVNDVTSAT 1992

Search completed: October 29, 2002, 09:31:00
Job time : 35.364 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:25:41 ; Search time 20.636 Seconds
(without alignments)
2556.865 Million cell updates/sec

Title: US-08-325-278b-1
Perfect score: 1565
Sequence: 1 AVENKEETPTDSEERV.....GGVTINIRFAGKKVDEKPEE 305

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTRMBL_19:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriaph:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1565	100.0	455	2 Q53291	Q53291 streptococc
2	1561	99.7	719	2 Q51912	Q51912 peptostrept
3	1226.5	78.4	992	2 Q51918	Q51918 peptostrept
4	167	10.7	664	2 Q53975	Q53975 streptococc
5	166	10.6	479	10 Q9LF88	Q9LF88 arabisopsis
6	166	10.6	669	2 Q93EM8	Q93EM8 streptococc
7	151	9.6	448	10 Q9SKP0	Q9SKP0 arabisopsis
8	149	9.5	448	10 Q96246	Q96246 arabisopsis
9	148.5	9.5	2276	2 Q93TV6	Q93TV6 staphylococ
10	141.5	9.0	1110	13 Q91255	Q91255 petromyzon
11	141	9.0	1849	2 Q9SK42	Q9SK42 lactobacill
12	137.5	8.8	992	16 Q9WZ46	Q9WZ46 thermotoga
13	137.5	8.8	5458	5 Q9U459	Q9U459 plasmodium
14	136	8.7	506	2 Q53837	Q53837 salmonella
15	134.5	8.6	3198	5 Q9U8G8	Q9U8G8 manduca sex
16	132.5	8.5	737	16 Q9JYK4	Q9JYK4 neisseria m

17	132.5	8.5	17352	5 Q95YM2	Q95YM2 procambus
18	131.5	8.4	891	2 Q04111	Q04111 enterococc
19	131.5	8.4	1166	2 Q86489	Q86489 staphylococ
20	130.5	8.3	584	2 Q69188	Q69188 streptococc
21	130.5	8.3	581	2 Q33742	Q33742 streptococc
22	130.5	8.3	663	2 Q30874	Q30874 streptococc
23	130.5	8.3	701	2 Q9KK48	Q9KK48 streptococc
24	130.5	8.3	1141	16 Q99W46	Q99W46 staphylococ
25	130.5	8.3	1141	16 Q932F7	Q932F7 staphylococ
26	129.5	8.3	1733	2 Q9K114	Q9K114 staphylococ
27	129	8.2	1171	2 Q9KW46	Q9KW46 staphylococ
28	128.5	8.2	938	16 Q927R4	Q927R4 listeria in
29	128.5	8.2	1119	13 P87344	P87344 theragra ch
30	127.5	8.1	701	2 Q9ROT5	Q9ROT5 streptococc
31	127.5	8.1	717	2 Q9EY85	Q9EY85 clostridium
32	127.5	8.1	728	2 Q9ZAN2	Q9ZAN2 neisseria g
33	127.5	8.1	843	2 Q47802	Q47802 enterococc
34	127	8.1	913	13 Q13099	Q13099 xenopus lae
35	126.5	8.1	1035	5 Q21380	Q21380 caenorhabdi
36	126.5	8.1	2045	16 Q9AOK5	Q9AOK5 streptococc
37	126	8.1	955	2 Q45574	Q45574 bacillus sp
38	125.5	8.0	673	10 Q9M9S5	Q9M9S5 arabisopsis
39	125.5	8.0	1510	5 Q25920	Q25920 plasmodium
40	125	8.0	463	10 Q39871	Q39871 glycine max
41	125	8.0	2083	5 Q9N435	Q9N435 caenorhabdi
42	124.5	8.0	1302	2 Q49547	Q49547 mycoplasma
43	124	7.9	495	16 Q99WG8	Q99WG8 staphylococ
44	124	7.9	506	2 Q53834	Q53834 salmonella
45	124	7.9	1489	10 Q96449	Q96449 phycophthor

ALIGNMENTS

RESULT 1

Q53291 ID Q53291 PRELIMINARY; PRT; 455 AA.

AC Q53291;

DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)

DE PROTEIN LG (FRAGMENT).

OS Streptococcus sp.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI_faxid=1306;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93094283; PubMed=1460053;

RA Kihlberg B.M., Sjöbrink U., Kastern W., Björck L.;

RT "Protein LG: a hybrid molecule with unique immunoglobulin binding properties.;"

RL J. Biol. Chem. 267:25583-25588(1992).

DR EMBL; S50809; AAA03280.1; -.

DR HSSP; P06654; 1PCX.

DR InterPro; IPR003147; B1.

DR InterPro; IPR00724; IgG_bind_B.

DR Pfam; PF02246; B1; 4.

DR Pfam; PF01378; IgG_binding_B; 2.

FT NON_TER 455 455

SQ SEQUENCE 455 AA; 49926 MW; 381FC235BBC8307B CRC64;

Query Match 100.0%; Score 1565; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 9.1e-79;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AVENKEETPTDSEERV...IKANLIFANGSTQAEFGTFEKATSEYAYADTLKKDN 60
|||||
Db 22 AVENKEETPTDSEERV...IKANLIFANGSTQAEFGTFEKATSEYAYADTLKKDN 81
|||||

Qy 61 GEYTVADVADKGYTLNFKAGKEKTPPEEKPEEVTIKANLIYADGKTQTAEFGTTEEPAAE 120
|||||

Db 82 GEYTVADVADKGYTLNFKAGKEKTPPEEKPEEVTIKANLIYADGKTQTAEFGTTEEPAAE 141
|||||

```
QY 121 AYRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTA 180
|||||
Db 142 AYRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTA 201
|||||
QY 181 EFKGTFEEATAEAYRYADLLAKENGYTVVDVADKGYTLNKFAGKEKTPPEPKKEVTIKA 240
|||||
Db 202 EFKGTFEEATAEAYRYADLLAKENGYTVVDVADKGYTLNKFAGKEKTPPEPKKEVTIKA 261
|||||
QY 241 NLIYADGKTQTAEFKGFPAEATAEAYRYADLLAKENGYTTADLEDGGYTTINIRFAGKKVD 300
|||||
Db 262 NLIYADGKTQTAEFKGFPAEATAEAYRYADLLAKENGYTTADLEDGGYTTINIRFAGKKVD 321
|||||
QY 301 EKPEE 305
|||||
Db 322 EKPEE 326
|||||

RESULT 2
Q51912
ID O51912 PRELIMINARY; PRT; 719 AA.
AC O51912;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PROTEIN L PRECURSOR.
OS Peptostreptococcus magnus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Finegoldia.
OX NCBI_TaxID=1260;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=312;
RX MEDLINE=90215984; PubMed=2108927;
RA Kastern W., Holst E., Nielsen E., Sjobring U., Bjorck L.;
RT "Protein L, a bacterial immunoglobulin-binding protein and possible
RT virulence determinant.";
RL Infect. Immun. 58:1217-1222(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=312;
RX MEDLINE=92316971; PubMed=1618782;
RA Bjoerck L., Sjoerbring U., Kastern W.;
RT "Structure of peptostreptococcal protein L and identification of
RT repeated immunoglobulin light chain-binding domain.";
RL J. Biol. Chem. 267:12820-12825(1992).
DR EMBL: M86697; AAA25612.1; -.
DR InterPro: IPR003147; B1.
DR Pfam: PF02246; B1; 5.
DR Pfam: PF00746; Gram_pos_anchor; 1.
KW Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 719 PROTEIN L.
SQ SEQUENCE 719 AA; 78983 MW; 963A8D76D5E34DD2 CRC64;

Query Match 99.7%; Score 1561; DB 2; Length 719;
Best Local Similarity 100.0%; Pred. No. 2.5e-78;
Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VENKEETPETDSEEVTIKANLIFANGSTQTAEFKGFEEKATSEAYAYADTLKKDNG 61
|||||
Db 95 VENKEETPETDSEEVTIKANLIFANGSTQTAEFKGFEEKATSEAYAYADTLKKDNG 154
|||||
QY 62 EYTVVDVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTAEFKGFEEKATSEAY 121
|||||
Db 155 EYTVVDVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTAEFKGFEEKATSEAY 214
|||||
QY 122 YRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTA 181
|||||
Db 215 YRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTA 274
|||||
QY 182 FKGTFFEATAEAYRYADLLAKENGYTVVDVADKGYTLNKFAGKEKTPPEPKKEVTIKAN 241
|||||
```

```
Db 275 FKGTFFEATAEAYRYADLLAKENGYTVVDVADKGYTLNKFAGKEKTPPEPKKEVTIKAN 334
|||||
QY 242 LIYADGKTQTAEFKGFPAEATAEAYRYADLLAKENGYTTADLEDGGYTTINIRFAGKKVDE 301
|||||
Db 335 LIYADGKTQTAEFKGFPAEATAEAYRYADLLAKENGYTTADLEDGGYTTINIRFAGKKVDE 394
|||||
QY 302 KPEE 305
|||||
Db 395 KPEE 398
|||||

RESULT 3
Q51918
ID O51918 PRELIMINARY; PRT; 992 AA.
AC O51918;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PROTEIN L PRECURSOR.
OS Peptostreptococcus magnus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Finegoldia.
OX NCBI_TaxID=1260;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3316;
RX MEDLINE=95078460; PubMed=7987012;
RA Murphy J.P., Trowern A.R., Duggleby C.J.;
RT "Nucleotide sequence of the gene for peptostreptococcal protein L.";
RL DNA Seq. 4:259-265(1994).
DR EMBL: L04466; AAA67503.1; -.
DR HSP: Q51911; IGAB.
DR InterPro: IPR003147; B1.
DR InterPro: IPR002988; GA.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF02246; B1; 4.
DR Pfam: PF01468; GA; 4.
DR Pfam: PF00746; Gram_pos_anchor; 1.
KW Signal.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 992 PROTEIN L.
SQ SEQUENCE 992 AA; 108700 MW; 9CFF5771578A5DCE CRC64;

Query Match 78.4%; Score 1236.5; DB 2; Length 992;
Best Local Similarity 79.9%; Pred. No. 8.5e-60;
Matches 246; Conservative 22; Mismatches 31; Indels 9; Gaps 4;

QY 2 VENKEETPETDSEEVTIKANLIFANGSTQTAEFKGFEEKATSEAYAYADTLKKDNG 61
|||||
Db 241 MERKLEKETPE--PEEVTIKANLIFADGSTQNAEFKGFKAQVSDAYAYADALKKNG 298
|||||
QY 62 EYTVVDVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTAEFKGFEEKATSEAY 121
|||||
Db 299 EYTVVDVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIFADGKTQTAEFKGFEEKATSEAY 358
|||||
QY 122 YRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTA 179
|||||
Db 359 YAYADLLAKENGYTTADLEDGGYTTINIRFAGKEKTPPEPKKEVTIKANLIYADGKTQTA 418
|||||
QY 180 AEPKGFEEATAEAYRYADLLAKENGYTVVDVADKGYTLNKFAGKEKTPPEPKKEVT 237
|||||
Db 419 AEPKGFEEATAEAYRYADLLAKENGYTTADLEDGGYTTINIRFAGKEKTPPEPKKEVT 478
|||||
QY 238 IKANLIYADGKTQTAEFKGFPAEATAEAYRYADLLAKENGYTTADLEDGGYTTINIRFAG 297
|||||
Db 479 IKVNLIFADGKTQTAEFKGFPAEATAEAYRYADLLAKENGYTTADLEDGGYTTINIRFAG 538
|||||
QY 298 KVDEKPEE 305
|||||
Db 539 ---EQPGE 543
|||||
```

```
RESULT 4
Q53975 PRELIMINARY; PRT; 664 AA.
AC Q53975;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DE IMMUNOGLOBIN G BINDING PROTEIN MIG PRECURSOR (IGG BINDING PROTEIN
DE MIG).
GN MIG.
OS Streptococcus dysgalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1334;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=SC1;
RX MEDLINE=94192673; PubMed=8143736;
RA Jonsson H., Mueller H.-P.;
RT "The type-III Fc receptor from Streptococcus dysgalactiae is also an
RT alpha 2-macroglobulin receptor.";
RL Eur. J. Biochem. 220:819-826(1994).
RN [2]
RP FUNCTION.
RX MEDLINE=99081771; PubMed=9864244;
RA Vasi J., Svensson J., Frick I.-M., Mueller H.-P.;
RT "Five homologous repeats of the protein G-related protein MIG
RT cooperate in binding to goat immunoglobulin G.";
RL Infect. Immun. 67:413-416(1999).
CC -1- FUNCTION: BINDS TO THE CONSTANT FC REGION OF IGG WITH HIGH
CC AFFINITY. N-TERMINUS MEDIATES BINDING TO PLASMA PROTEINASE
CC INHIBITOR ALPHA 2-MACROGLOBULIN AFTER COMPLEX FORMATION WITH
CC PROTEASES.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL (BY
CC SIMILARITY).
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS IN
CC THE REGION OF THE MEMBRANE ANCHOR. C-TERMINAL IS SIMILAR TO OTHER
CC STREPTOCOCCAL TYPE-III FC RECEPTORS.
DR EMBL: Z29666; CAA82764.1;
DR HSSP: P06654; 1PGX
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR000724; Igg_bind_B.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR Pfam: PF01378; Igg_binding_B; 5.
DR PRINTS: PR00015; GP0SANCHOR
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
KW Igg-binding protein; Repeat; Transmembrane; Cell wall; Signal.
FT SIGNAL 1 30
FT CHAIN 31 664
FT DOMAIN 31 208
FT DOMAIN 559 638
FT DOMAIN 639 659
FT TRANSMEM 660 664
FT DOMAIN 660 664
FT DOMAIN 87 189
FT DOMAIN 87 110
FT REPEAT 166 189
FT REPEAT 209 558
FT DOMAIN 209 558
FT REPEAT 209 278
FT REPEAT 279 348
FT REPEAT 349 418
FT REPEAT 419 488
FT REPEAT 489 558
FT DOMAIN 592 621
FT DOMAIN 630 635
FT SEQUENCE 664 AA; 72023 MW; 4A4ICEEF7977862A CRC64;
Query Match 10.7%; Score 167; DB 2; Length 664;
Best Local Similarity 27.4%; Pred. No. 0.08;
```

```
Matches 84; Conservative 49; Mismatches 134; Indels 40; Gaps 16;
Qy 7 ETPTPTDSEEEVTKANLIFANGSTQTAEFKGTFEKATSEAYAYADTLKKDNGEYTD 66
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
214 DAPELTPALTYKLVVKN-TFSGETTTKADTATAKEFKQ-YATANNV---DGEWSYD 268
Qy 67 VADKGYTLNKA---GKEKTPPEPKKEVTKANLIYADGKTQTAEFKGFEEATAYR 123
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
269 DATKTFVTEKPAVIDAPELTPALTYKLVKGNF--SGETTT--KAVDATAEAKAFK 323
Qy 124 -YADALKKDNGETYVDVADKGYTLNKA---GKEKTPPEPKKEVTKANLIYADGKTQ 179
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
324 QYATANNVD--GEMSYDDATKTFVTEKPAVIDAPELTPALTYKLVKGNF--SGETTT 380
Qy 180 AEFKGTFEETAEAYRYADLLAKENGKTYVDVADKGYTLNKA---GKEKTPPEPKKE 236
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
381 KADAATAEKEFKOYATANGV---DGEWSYDDATKTFVTEKPAVIDAPELTPALTYK 437
Qy 237 TIKANLIYADGKTQTAEFKGTFEATAEAYRYADLLAKENGKTYADLEDGGYTNIRFAG 296
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
438 IVKGNF--SGETTT--KAVDATAEAKAFKQ---YANENGVY-----GEMSYDDATKT 483
Qy 297 KKVDEKP 303
Db : : : :
484 FTVTEKP 490
RESULT 5
Q9LF88 PRELIMINARY; PRT; 479 AA.
AC Q9LF88;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE LATE EMBRYOGENESIS ABUNDANT PROTEIN-LIKE.
GN F8J2.210.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nyakatura G., Partmann B., Dauner D., Sterr W., Holland R.,
RA Weichselgartner M., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.,
RA Quettier F., Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsids sequencing project;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL132969; CAB86908.1;
DR InterPro: IPR004238; LEA.
DR Pfam: PF02987; LEA; 5.
SQ SEQUENCE 479 AA; 52084 MW; C39E00D910E2385 CRC64;
Query Match 10.6%; Score 166; DB 10; Length 479;
Best Local Similarity 25.8%; Pred. No. 0.062;
Matches 82; Conservative 43; Mismatches 137; Indels 56; Gaps 12;
Qy 3 ENKETPTPTDSEEEVTKANLIFANGSTQTAEFKGTFEKATSEAYAY----ADTLKK 58
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
75 DTAESTREGADIASEKAAGMR-----DTTGEVRDSTAQKTETADYADKAREAKD 125
Qy 59 DNGEYTVDVADKGYTLNKAIFAGKTPPEPKKEVTKANLIYADGKTQTAEFKGTFEAT 118
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
126 KTADKTETAD--YAEAKAREAKDRTADTKETAETAEKA-REAKOKTADKLGKDYD 182
Qy 119 AEAYRYA-DALKKDNGETYVDVADKGYTLNKAIFAGKTPPEPKKEVTKANLIYADGKT 177
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
183 AEAKAREAKDTAEKLGKY-----KDYTVDKAKAKDKTAETAEYTSDKA-RETKD 235
Qy 178 QTAEFKGTFEETAE-----AYRYADLLAKENGKTYVDVADKGYTLNKAIFAGK 229
```

```

Db 236 KTAEEKVEYDYTAERAKADTAERAKDVEYDYTAEK--ATETKDAGVSKIG 293
QY 230 EEPKEEV-TIKANLIYADGKTQ-----TAEFKGTFAEATAEAYRADLLAKENGKGTADL 283
Db 294 ELKDSAVDTAKRAMGLUSGKTEETKOKAVETKDTAKEKMDKEAGEEARRKWE----- 345
QY 284 EDGGYTINIRFAGKKVDE 301
Db 346 -----MRLEGGKKLDE 355

RESULT 6
Q93EM8 PRELIMINARY; PRT; 669 AA.
AC Q93EM8;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DE MIG PRECURSOR.
OS Streptococcus dysgalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 43078;
RX MEDLINE=21437603; PubMed=11553540;
RA Song X.M., Perez-Casal J., Bolton A., Potter A.A.;
RT "Surface-Expressed Mig Protein Protects Streptococcus dysgalactiae
RT against Phagocytosis by Bovine Neutrophils.";
RL Infect. Immun. 69:6030-6037(2001).
DR EMBL; AF354651; AAL09476.1; -.
KW Signal.
FT SIGNAL.
FT CHAIN.
FT CHAIN.
SQ SEQUENCE 669 AA; 72682 MW; 5C8982B952029B03 CRC64;

Query Match 10.6%; Score 166; DB 2; Length 669;
Best Local Similarity 27.5%; Pred. No. 0.091;
Matches 85; Conservative 48; Mismatches 132; Indels 44; Gaps 16;

QY 7 ETPTPTDSEEVTIKANLIIFANGSTQTAEFKGTPEKATSEAYAYADTLKNGEYTV 66
Db 214 DAPELTPALTYKLIVKGN-TFSGETTTKAIDTATAEKFQ-YATANNV--DGEWSYD 268
QY 67 VADKGYTLNKFKA--GKEKTPPEKPEEVIKANLIYADGKTQTAEFKGTPEEATAEA-- 121
Db 269 DATKFTFTVEKPAVIDALELTPALTYKLIVKGNTE--SGETTTPK----AIDAATAEFK 322
QY 122 YRYADALKKNGEYTVVADKGYTLNKFKA--GKEKTPPEKPEEVIKANLIYADGKTQ 178
Db 323 KOYATANNV-DGEWSYDYATKFTFTVEKPAVIDAPELTPALTYKLIVKGNTE--SGETT 379
QY 179 TAEFKGTFAEATAEAYRADLLAKENGKGYTVVADKGYTLNKFKA--GKEKTPPEKPEE 235
Db 380 TKALDAATAEKEFKOYATAN--NVDEWSYDDATKFTFTVEKPAVIDAPELTPALTYK 436
QY 236 VTIKANLIYADGKTQTAEFKGTFAEATAEAYR-YADLLAKENGKGYTADLEGGYTINIRF 294
Db 437 LIVKGNFTF--SGETTTPK--KAVDAETAERAKFOYA-----TANNVDGEWSYDDAT 481
QY 295 AGKKVDEKP 303
Db 482 KFTVTEKP 490

RESULT 7
Q9SKP0 PRELIMINARY; PRT; 448 AA.
AC Q9SKP0;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)

```

```

DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE LATE EMBRYOGENESIS ABUNDANT PROTEIN (ATECP63).
GN AT2G36640.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OX eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umavam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AC006282; AAD20140.1; -.
DR InterPro; IPR004238; LEA.
DR Pfam; PF02987; LEA; 4.
SQ SEQUENCE 448 AA; 48492 MW; C72563D4194DD9BF CRC64;

Query Match 9.6%; Score 151; DB 10; Length 448;
Best Local Similarity 24.5%; Pred. No. 0.38;
Matches 79; Conservative 38; Mismatches 107; Indels 98; Gaps 14;

QY 3 ENKEETPTDSEEVTIKANLIIFANGSTQTAEFKGTPEKATSEAYAYADTLKNDGE 62
Db 79 EAAESTKEGAQIASEKAVGAK-----DATVEKAKET----ADVTAEKVGE 119
QY 63 YIVVDADKGYTLNIRFAGKTEPPEKPEEVIKANLIYADGKTQTAEFKGTPEEATAEAY 122
Db 120 Y-----KDYTVDKAKEAKDTTAEKAKETANYTADKA-VEAKDKTAEKIGEYK----- 165
QY 123 RVADALKKNGEYTVVADKGYTLNIRFAGKTEPPEKPEEVIKANLIYADGKTQTAEF 182
Db 166 -----DYAVDKA-----VEAKDKTAEKAKETANYTADKA-KEAKDKTAEK 204
QY 183 KGTFFEATAE-----AYRYADLLAKENGKGYTVVADKGYTLNIRFAGKTEPTE 230
Db 205 VGEYKDYTVDKAVEARDYTAERAKAIAEAKDKTAEKTEG-----KDYTVKATEGKDVTS 258
QY 231 ---EPKEEV--TIKANLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGKGYTADLE 285
Db 259 KIGELKDSAVEAKRAMGLSKTEEAKGKAVETKDT-----AKEN-----MEK 302
QY 286 GGVTI-----NIRFAGKKVDEK 302
Db 303 AGEVTRQKWEEMRLEGGKELKEE 324

RESULT 8
Q96246 PRELIMINARY; PRT; 448 AA.
AC Q96246;
DT 01-FEB-1997 (T-EMBLrel. 02, Created)
DT 01-FEB-1997 (T-EMBLrel. 02, Last sequence update)
DE LEA PROTEIN IN GROUP 3.
GN ATECP63.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

```


OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;

RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA; TISSUE=DRY SEED;

RX MEDLINE=97169149; PubMed=9016956;

RA Yang H., Saitou T., Komeda Y., Harada H., Kamada H.;

RT "Arabidopsis thaliana ECP63 encoding a LEA protein is located in

RT chromosome 4.";

RL Gene 184:83-88(1997).

DR EMBL; D64140; BAA11017.1; -.

DR InterPro; IPR004238; LEA.

DR Pfam; PF02987; LEA; 4.

SQ SEQUENCE 448 AA; 48534 MW; DC370999F18FBA85E CRC64;

Query Match 9.5%; Score 149; DB 10; Length 448;

Best Local Similarity 24.5%; Pred. No. 0.49;

Matches 79; Conservative 38; Mismatches 107; Indels 98; Gaps 14;

QY 3 ENKEETPETDSEEVITKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKDNGE 62

Db 79 EAESTKEGAQIASEKAVGAK-----DATVEKAKET----ADYAEKVGE 119

QY 63 YTVDAVKGTYLNTKFKAGKTPPEKPEVITKANLIYADGKTQTAEFKGTPEATAEAY 122

Db 120 Y-----KDYTVDAKAEKDTTAEKAKETANYTADKA-VEAKDKTAEKIGEK----- 165

QY 123 RYADALKDNGEYTVDAVKGTYLNTKFKAGKTPPEKPEVITKANLIYADGKTQTAEF 182

Db 166 -----DYAVDKA-----VEAKDKTAEKAKETSNTADKA-KEAKDKTAEK 204

QY 183 KGTPEATAE-----AYRYADLLAKENGKTYTVDAVKGTYLNTKFKAGKTPPE 230

Db 205 VGEKDYTVDAKAVARDYTAEKAEAKDKTAEKIGY-----KDYVEKATEGKDYTVS 258

QY 231 ---EPKEEV---TIKANLIYADGKTQTAEFKGTPEATAEAYRYADLLAKENGKTYADLE 285

Db 259 KLGLKDSAVETAKRAMGFLSGKTEAKGAVETKDT-----AKEN-----MEK 302

QY 286 GGYTI-----NIRPAGKVDK 302

Db 303 AGEVTRQKMEEMRLGKELKEE 324

RESULT 9
Q93TY6

ID Q93TY6 PRELIMINARY; PRT; 2276 AA.

AC Q93TY6;

DT 01-DEC-2001 (Tremblrel. 19, Created)

DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE BIOFILM-ASSOCIATED SURFACE PROTEIN.

GN BAP.

OS Staphylococcus aureus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Staphylococcus.

OX NCBI_TaxID=1280;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=V329;

RX MEDLINE=21189316; PubMed=11292810;

RA Cucarella C., Solano C., Valle J., Amorena B., Lasa I.;

RA Penades J.R.;

RT "Bap, a Staphylococcus aureus Surface Protein Involved in Biofilm

RT Formation.";

RL J. Bacteriol. 183:2888-2896(2001).

DR EMBL; AF288402; AAK38834.1; -.

SQ SEQUENCE 2276 AA; 238792 MW; 8623C60472CB0C0F CRC64;

Query Match

Best Local Similarity 9.5%; Score 148.5; DB 2; Length 2276;

Matches 72; Conservative 36; Mismatches 135; Indels 97; Gaps 15;

QY 8 TPETPTDSEEVITKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKNGEYTVDV 67

Db 1644 TSDATQVTGQAEPNSTVKLFPDGTAT-----GT-----AD-----DQNYTIDI 1684

QY 68 ADK-----GYTLNKFAGKTKTPEPK-----EVTIKANLIYADGKTQTF 107

Db 1685 PSNVDLNGGELQVATDKDGTSEPSANVTDTTAPDAPTVNDVTSATQVTGQAEPS 1744

QY 108 AEFKGTPEEATAEAYRYADALKKNGEYTVDVADK-----GYTLNKFAGKTKTPEPK- 161

Db 1745 T-VKLTFPDGTT-----ATGTADDOGNYTIDIPSNVLDNGGELQVATDKDGTSEPKL 1798

QY 162 -----EVTIKANLIYADGKTQTAEFKGTPEEATAEAYRYADLLAKENGY 207

Db 1799 TNVTDTTASDAPTVNDVTSASQVTGQAEPNST-VKLTFPDGTT-----ATGTADDOGNY 1852

QY 208 TVDVADK-----GYTLNKFAGKTKTPEPK-----EVTIKANLIYADG 247

Db 1853 TIDIPSNVLDNGGELQVATDKDGTSEPSANVTDTTAPDAPTVNDVTSATQVTGQA 1912

QY 248 KTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEDGG 287

Db 1913 EPNST-VKLTFPDGTT-----ATGTADDOGNYTIDIPSNVLDNGGELQVATDKDGTSEPKL 1946

RESULT 10
Q91255

ID Q91255 PRELIMINARY; PRT; 1110 AA.

AC Q91255;

DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE NF-180.

OS Petromyzon marinus (Sea lamprey).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;

OC Petromyzontiformes; Petromyzontidae; Petromyzon.

OX NCBI_TaxID=7757;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=CNS;

RX MEDLINE=95287814; PubMed=7770000;

RA Jacobs A.J., Kamholz J., Selzer M.E.;

RT "The single lamprey neurofilament subunit (NF-180) lacks

RT multiphosphorylation repeats and is expressed selectively in

RT projection neurons.";

RL Brain Res. Mol. Brain Res. 29:43-52(1995).

DR EMBL; U19361; AAA80106.1; -.

DR InterPro; IPR001664; IF.

DR InterPro; IPR000533; Tropomyosin.

DR Pfam; PF00038; filament; 1.

DR PRINTS; PR00194; TROPOMYOSIN.

DR PROSITE; PS00226; IF; UNKNOWN_1.

SQ SEQUENCE 1110 AA; 123818 MW; 6558DA73DAF6974C CRC64;

Query Match 9.0%; Score 141.5; DB 13; Length 1110;

Best Local Similarity 25.4%; Pred. No. 3.6;

Matches 87; Conservative 38; Mismatches 138; Indels 79; Gaps 13;

QY 1 AVENKEETPETPDSEEVITKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKD 59

Db 482 SAEKDEEEEEEEKEEEEE-----AAEEEEEDRGRKEGEAAEAEAEVEEKEE 532

QY 60 NGEYTVDVADKGYT-----LNKFKAGKTKTPEPKPEVITKANLIYADGKT 105

Db 533 AEEAE 584

QY 106 QTAEFKGTPEEATAE 156

Db 585 EAAAKAE 638

QY 157 PEE--PKPEVITKANLIYADGKTQTAEFKGTPEEATAEAYRYADLLAKENGYTVDVADK 214

Db 639 EEEEEAEAEVTSK-----KAKTQEAEE-----VEEEAEAEAEAEAEAEAEAE 684

QY	215	GYTLNTKFRAGKCKTPEPK-----EVTIKANLIYADGKTQTAFRTGTTAAATAE	264
Db	685	-----VAASKEEEDSDKEADAEDEAEVEEVKSEEVTKSDAEAEAEAEAAKSEEE	739
QY	265	AYRYA-DLLAKENGKYTALEDDGGYTNIRFAGKKVDKPEE	305
Db	740	AAEEAKDEAEAEAEAEVVEETAATEAEAEASDDCKPEE	781

```

RESULT 11
Q9S4K2 PRELIMINARY; PRT; 1849 AA.
ID
Q9S4K2;
AC
DT 01-WAY-2000 (TREMBLrel. 13, Created)
DT 01-WAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELL ENVELOPE-ASSOCIATED PROTEINASE.
GN PRTH.
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
NCBI_Taxid=1587;
OX
RN {1}
SEQUENCE FROM N.A.
RP
RC STRAIN=CNR232;
RC MEDLINE=99350424; PubMed=10419958;
RA Pederson J.A., Mileski G.J., Weiner B.C., Steele J.L.;
RT "Genetic characterization of a cell envelope-associated proteinase
RT from Lactobacillus helveticus CNR232.";
RL J. Bacteriol. 181:4592-4597(1999).
DR EMBL; AF133727; AAD50643.1; -
DR HSP; P00782; 1SUC.
DR MEROPS; S08.018; -.
DR InterPro; IPR001444; Flag_bb_rod.
DR InterPro; IPR003137; PA.
DR InterPro; IPR00209; Peptidase_S8.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 3.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
DR PROSITE; PS00136; SUBTILASE_ASF; UNKNOWN_1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 1849 AA; 199614 MW; 219F0D44B15A091F CRC64;

Query Match 9.0%; Score 141; DB 2; Length 1849;
Best Local Similarity 27.7%; Pred. No. 6.9;
Matches 76; Conservative 32; Mismatches 108; Indels 58; Gaps

```

Qy	73	TLN	IKFAGKEKTPPEEK	EEVT	IKAN	LIYADG	KTQTAEPK	GTPEAT	A6A	--	YRYADAL	K 130																							
		:			:	:	:																												
Db	1415	TLNL	---DSEN	VYVYNNKDKFTI	-SGT	ISDDYK	FYDLS	INGNDV	ETWSA	VYDYS	HSKEG	K 1470																							
		:		:	:	:	:																												
Qy	131	DNGEY	TVDVAD	KG	YTLN	IK	-----	---FAG	KEKTPPE	KPEE	--	VTKAN	LIYA 173																						
		:		:	:	:	:																												
Db	1471	-NPK	HEVD	LKG	KN	TFNV	KVTD	IQGN	SSSQAL	VVY	YEP	AKT	LAEP	SV	D	KLL	TK	TAN	LQ	L 1529															
		:		:	:	:	:																												
Qy	174	DGKT	QTAE	FK	-----	---GT	PEATA	EAY	RYAD	LL	LAKEN	GYTV	DV	AD	KG	Y	-----	---TL	218																
		:		:	:	:	:																												
Db	1530	KAPT	DESE	AK	VY	SLD	NG	KT	FND	VP	AD	GFK	-----	---VT	ENG	TVQ	VP	K	AD	K	YGNES	KV	S 1583												
		:		:	:	:	:																												
Qy	219	NIK	FAG	KEK	TPPE	E	EV	TI	KAN	L	---IY	AD	KG	TQ	---	---TAE	FG	TAE	A	T	AY	RYAD	L 271												
		:		:	:	:	:																												
Db	1584	EIK	L	KN	E	KNQ	P	SE	D	E	LAK	E	N	LQ	AK	Y	DAGE	K	D	L	D	K	Y	TAD	S	K	F	D	N	AL	K	A	---	KDV	1640
		:		:	:	:	:																												
Qy	272	LAKEN	GY	TAD	LE	D	GG	Y	T	I	N	R	FAG	K	V	DE	K	P	E	E	305														
		:		:	:	:	:																												
Db	1641	LAD	KN	AK	-LAD	LQ	DA	AK	ALD	-	KAP	QAL	TE	K	P	AE	1671																		

RESULT 12
Q9WZA6

ID Q9WZA6 PRELIMINARY; PRT; 992 AA.
 AC Q9WZA6;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE POLYSACCHARIDE EXPORT PROTEIN, PUTATIVE.
 GN TM0638.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 OC NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=9287316; PubMed=10360571;
 RA NELSON K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 McDonald L., Uterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 genome sequence of Thermotoga maritima";
 RL Nature 399:323-329(1999).
 DR EMBL; AE001738; AAD35722.1; -.
 DR TIGR; TM0638; -.
 KW Complete proteome.
 SQ SEQUENCE 992 AA; 109623 MW; 19291DDF29721BF3 CRC64;

 Query Match 8.8%; Score 137.5; DB 16; Length 992;
 Best Local Similarity 22.3%; Pred. No. 5.3;
 Matches 77; Conservative 40; Mismatches 112; Indels 117; Gaps

 QY 2 VENKEETPEPTDSREEVTIKANLIFANGSTQAE-----FKGTPEKANS 47
 Db 219 VLSKDETFVSEKDIYQ-----NLLKPGDTVHVQKIDRFAYVGVAVRPGMYTFSRE 273
 QY 48 EAYADATLKKDNG-----EYTDVADKGYTLNI-----KFAQG 80
 Db 274 ESLTLKLNIAKAGGVSDKRFIEKALITRGGETLEYDPOVLDENVEIEGVDVVEIKKF-- 331
 QY 81 KEKTP-----EPKEEVIKANLIYADKGTQTAEFKGTFFEATAEAYRYAD 126
 Db 332 -EKTEVYGVSVRPGVYEISPRENVTLKLLSMAGG-----IKGTLEVD-----SI 378
 QY 127 ALKDCNGEYTDVADKGYTLNIFAGKEKTPPEEKVEITKANLIYADKGTQTAEFKGTFF 186
 Db 379 VTRDGSVTLS-----PNLDFSVKPGDVNVKFEVPKKA---YVLGVVRN---PGLY 426
 QY 187 EEATAEAYRYADLLAKENGRTYTDVADKGYTLNIFAGKEKTPPEEKVEITKANLIYAD 246
 Db 427 TFGKNEAFTLRNLIAKAGG-----FYDEGQVSVKVGAKKEYSPDE-----IVKED 476
 QY 247 GKQTQAEFKGTFFEATAEAYRYADLL-----AKENGKYTADLED 285
 Db 477 GVFEYVE-----RYTDRFVYVWGDVNSRNGKMSFEKEE 509

 RESULT 13
 Q9U459 PRELIMINARY; PRT; 5458 AA.
 AC Q9U459;
 ID Q9U459 Q9GPG4;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE ERYTHROCYTE MEMBRANE-ASSOCIATED GIANT PROTEIN ANTIGEN 332.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 OC NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FC01/JHN;
 RA Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M., Lu J.H., X J.;
 RT "Molecular cloning and structure analysis of the Plasmodium falciparum

Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M., Lu J.H., X J.;
Strain=FCCI/HN;
"Molecular cloning and structure analysis of the Plasmodium falciparum

```
RT erythrocyte membrane-associated giant protein Aq332 (PF332) gene." ;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF202180; AAF15293.3; -
DR InterPro: IPR001313; PUM.
SQ SEQUENCE 5458 AA; 615269 MW; 4F6A37D92BD80172 CRC64;

Query Match      8.8%; Score 137.5; DB 5; Length 5458;
Best Local Similarity 20.5%; Pred. No. 37;
Matches 67; Conservative 55; Mismatches 138; Indels 67; Gaps 12;

QY 3 ENKEET---PETPETSEEEVTTKANLI-FANGSTQTAEF-----KGTFEKATSBAYAYADT 55
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1928 EKSEESAIEBEVKEVKEEISEK---IVEKEESVTEIEVRQBEESTTEKIVKDVSPTEDF 1984
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 56 LKK-----DNGEYTVDVADKGYTLNKKFAGKKEKTPPEEKVEVTTKANLIYAD 102
   :|:|  :|:|  :|:|  :|:|  :|:|  :|:|  :|:|  :|:|  :|:|  :|:|  :|:|  :|:|
Db 1985 VEQDTSVTEKIVQEGSNTAEVDVEK-----ESASDEHEQEDSVNAQV----- 2030
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 103 GKQTQAEFKTFEEAEAYADALKKNGEYTVDVADKGYTLNKKFAGKKEKTPPEPK 162
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2031 ---TCEKKSVTREIVDEVS-RTEEIVEENGSVTEGVDETG-----SVTEEIIE 2074
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 163 EVTIKANLIYADGKTQTAEFKGFEEATAEAYRYADLLAKENGKYTVDVADKGYTLNKF 222
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2075 EATVTEEVVEDGSVTERVEDGVIQEVVEDGSVTEIEVQENGSVTEIEVEEGSVN--- 2131

QY 223 AGKEKTPPEEKVEVTTKANLIYADGKTQTAEFKGFEEATAEAY-RYADLLAKENGKYTA 281
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2132 -----EEVEEVSVEGGEVDETYVTEVEEVEEGSVVVEEVEEGSVVVEEVEEGSVVE 2184
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 282 DL-EDGGYTIN--IRFAGKKVDEKPEE 305
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2185 EIVEEGSVVVEEVEEGSVVVEEIVEE 2211

RESULT 14
Q53837 PRELIMINARY; PRT; 506 AA.
ID Q53837
AC Q53837
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PHASE-2 FLAGELLIN STRUCTURAL PROTEIN.
GN FLJJB.
OS Salmonella cholerae-suis (Salmonella enterica).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=591;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN=SEROTYPE CHOLERAESUIS, ATCC 7001;
RX MEDLINE=95325331; PubMed=7541401;
RA Vanegas R.A., Joys T.M.;
RT "Molecular analyses of the phase-2 antigen complex 1,2. . of Salmonella spp." ;
RL J. Bacteriol. 177:3863-3864(1995).
DR EMBL: U17175; AAC43352.1; -
DR InterPro: IPR001029; Flagellin_C.
DR InterPro: IPR001492; Flagellin_N.
DR Pfam: PF00700; Flagellin_C; 1.
DR Pfam: PF00669; Flagellin_N; 1.
DR PRINTS: PR00207; FLAGELLIN.
DR ProDom: PD000316; Flagellin_C; 1.
SQ SEQUENCE 506 AA; 52590 MW; B9E41977F0A96A20 CRC64;

Query Match      8.7%; Score 136; DB 2; Length 506;
Best Local Similarity 24.4%; Pred. No. 2.9;
Matches 94; Conservative 44; Mismatches 129; Indels 118; Gaps 21;

QY 1 AVENKETPETPETDS--EEVEVTTKANLI-FANGSTQTAEFKGFEEKATSBAYAYADTLKK 58
   |||:|  |||:|  |||:|  |||:|  |||:|  |||:|  |||:|  |||:|  |||:|  |||:|
Db 96 AVOSANSTNSQSDLSQAEITQRLNEIDRVSGTQ---FNGV--KVLQADNTLTQVGA 150
```

```
QY 59 DNGEYTVDVADK---GYTLNKKFAGKKEKTPPEEKVEVTTKANLIYAD----- 102
   :|:|  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 151 NDGE-TIDIDLKQINSQTLGLDSLNQKAYDKVDTAVTTKA---YANNGTTLDSVGLDDT 206
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 103 -----GKTQTAEFKGFEEATAEAYRY-----ADALKKNGEYTVDVADKGYTL 146
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 207 AIKAAIGGTTGTAPVTGSAVKFADNNKRYFVTIGGFTGADAAG--NGDYEYVNAVTDG-TV 263
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 147 NIKFAGKKEKTPPEEKVEVTTKANLIYADGKTQTAEFKGFEEATAEAE----- 193
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 264 TLA-AGATKT-----TMPAG---ATTTEVOELKDTPAVVSADAKNALIAGGVDA 311
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 194 -----YRYADLLAKE-NGKYTVDVADKGYTLNKF-----GKEKTP 229
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 312 ANGAELVKMSYTDKNGKTIDGGYALKAGDYAAADYDEATGAIKAKTTSYTAADGTTKTA 371
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 230 EEPKEVTTKANLIYADGKTQTA-----BFGK--TFEATAE-----AYYAD 270
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 372 ANOLGGVDGKTEVVTIDGKTYNASKAAGHDFKAQPELAEEAAKTTPNLOKIDAALQVD 431
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 271 LLAKENG---KYTADLEDGGYTIN 291
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 432 ALRSDLGAVQNRFSNATNLGNTVN 456

RESULT 15
Q9U8G8 PRELIMINARY; PRT; 3198 AA.
ID Q9U8G8
AC Q9U8G8
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE LACUNIN PRECURSOR.
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Sphingioidea; Sphingidae; Sphinginae; Manduca.
OX NCBI_TaxID=7130;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99457716; PubMed=10528409;
RA Nardi J.B., Martos R., Walden K.K., Lampe D.J., Robertson H.M.;
RT "Expression of lacunin, a large multidomain extracellular matrix
   protein, accompanies morphogenesis of epithelial monolayers in Manduca
   sexta." ;
RL Insect Biochem. Mol. Biol. 29:883-897(1999).
DR EMBL: AF078161; AAF04457.1; -
DR HSSP: P12111; 2KNT.
DR InterPro: IPR004094; Antistasin.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR002223; Kunitz_BPTI.
DR InterPro: IPR000884; TSPI.
DR InterPro: IPR002221; WAP.
DR Pfam: PF02822; Antistasin; 4.
DR Pfam: PF00047; Ig; 2.
DR Pfam: PF00014; Kunitz_BPTI; 9.
DR Pfam: PF00095; wap; 1.
DR PRINTS: PR00759; BASICPTASE.
DR SMART: SM00408; IGC2; 2.
DR SMART: SM00131; KU; 10.
DR SMART: SM00209; TSPI; 7.
DR SMART: SM00217; WAP; 1.
DR PROSITE: PS00317; 4-DISULFIDE_CORE; 1.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 8.
DR PROSITE: PS00279; BPTI_KUNITZ_2; 10.
DR PROSITE: PS00092; TSPI; 1.
KW Immunoglobulin domain; Serine protease inhibitor; Signal.
FT SIGNAL 1 21 POTENTIAL.
SQ SEQUENCE 3198 AA; 349364 MW; AB4ACD459C0D9134 CRC64;

Query Match      8.6%; Score 134.5; DB 5; Length 3198;
Best Local Similarity 24.8%; Pred. No. 30;
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:25:26 ; Search time 7.42896 Seconds
(without alignments)
1589.653 Million cell updates/sec

Title: US-08-325-278b-1
Perfect score: 1565
Sequence: 1 AVENKEETPETDSEEV.....GGYTINIRFACKKVDEKPEE 305

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	151.5	9.7	183	1 CNA_STAAU	Q53654 staphylococ
2	141	9.0	555	1 EDC8_DAUCA	P20075 daucus caro
3	130.5	8.3	1337	1 DEXT_STRDO	P39653 streptococc
4	129	8.2	505	1 FLJB_SALTY	P52616 salmonella
5	127.5	8.1	1940	1 MYH3_CHICK	P02565 gallus gall
6	127	8.1	2459	1 MAPB_RAT	P15205 rattus norv
7	125	8.0	2468	1 MAPB_HUMAN	P46821 homo sapien
8	123.5	7.9	2464	1 MAPB_MOUSE	P14873 mus musculu
9	122	7.8	332	1 SP2B_BACSU	P37575 bacillus su
10	122	7.8	593	1 SPG2_STRSP	P19909 streptococc
11	122	7.8	1196	1 ICEV_PSEIX	O33479 pseudomonas
12	122	7.8	1200	1 ICEN_PSESY	P06620 pseudomonas
13	121.5	7.8	1500	1 SSF5_STRGN	P16952 streptococc
14	121.5	7.8	1637	1 MRSP_STAAU	P80544 staphylococ
15	121	7.7	627	1 ABPX_YEAST	Q08641 saccharomyc
16	119.5	7.6	848	1 NFM_MOUSE	P08553 mus musculu
17	117.5	7.5	650	1 NUCI_XENLA	P20397 xenopus lae
18	117.5	7.5	845	1 NFM_RAT	P12397 rattus norv
19	117.5	7.5	1435	1 SUBF_BACSU	P16397 bacillus su
20	117	7.4	1148	1 ICEK_PSEIX	O30611 pseudomonas
21	116.5	7.4	694	1 NUCI_CHICK	P15771 gallus gall
22	116.5	7.4	1935	1 MYSS_CYPCA	Q90339 cyprinus ca
23	116	7.4	915	1 NFM_HUMAN	P07197 homo sapien
24	115.5	7.4	793	1 YF06_MYCPN	P75280 mycoplasma
25	115	7.3	272	1 OSAT_BORBU	Q04968 borrelia bu
26	114.5	7.3	500	1 FLJB_SALAE	P52615 salmonella
27	114.5	7.3	1394	1 HAP_HAEIN	P45387 haemophilus
28	114.5	7.3	1409	1 HAP1_HAEIN	P44596 haemophilus
29	114	7.3	2116	1 MY52_DICDI	P08799 dictyosteli
30	112.5	7.2	300	1 SIAP_PIG	P31936 sus scrofa
31	112.5	7.2	1004	1 SLPO_BACBR	P09333 bacillus br
32	112.5	7.2	1104	1 COLA_CLOPE	P43153 clostridium
33	111.5	7.1	424	1 IRK8_MOUSE	P97794 mus musculu

RESULT 1

ID	CNA_STAAU	STANDARD;	PRT;	1183 AA.
AC	Q53654:			
DT	15-DEC-1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Collagen adhesin precursor.			
GN	CNA.			
OS	Staphylococcus aureus.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Staphylococcus.			
OX	NCBI_TaxID=1280;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FDA 574;			
RC	MEDLINE=92165839; PubMed=1311320;			
RA	Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,			
RA	Lindberg M., Hoeoek M.;			
RL	J. Biol. Chem. 269:11672-11672(1994).			
[3]				
RP	COLLAGEN-BINDING DOMAIN.			
RC	STRAIN=FDA 574;			
RC	MEDLINE=94032261; PubMed=8218209;			
RA	Patti J.M., Bolles J.O., Hoeoek M.;			
RA	"Identification and biochemical characterization of the ligand			
RT	binding domain of the collagen adhesin from Staphylococcus aureus."			
RL	Biochemistry 32:11428-11435(1993).			
[4]				
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.			
RA	MEDLINE=97475225; PubMed=9334749;			
RA	Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,			
RA	Moore D., Jin L., Schneider A., DeLucas L.J., Hoeoek M.,			
RA	Narayana S.V.L.;			
RT	"Structure of the collagen-binding domain from a Staphylococcus			
RT	aureus adhesin.";			
RL	Nat. Struct. Biol. 4:833-838(1997).			
CC	-!- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO			
CC	COLLAGEN-CONTAINING SUBSTRATA.			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.			
CC	-!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS			
CC	IN THE REGION OF THE MEMBRANE ANCHOR.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			

P43597 saccharomyc
Q54114 streptococc
P48467 neurospora
Q53957 streptococc
Q62234 mus musculu
P12882 homo sapien
P05423 homo sapien
P16620 drosophila
P06176 salmonella
P13466 dictyostell
Q28641 oryctolagus
P21333 homo sapien

ALIGNMENTS

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M81736; AAA20874.1; -.
DR PDB; 1AMX; 24-JUN-98.
DR InterPro; IPR001899; Gram_pos_anchor.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
KW Signal; Repeat; Transmembrane; Cell wall; 3D-structure.
FT SIGNAL 1 29
FT CHAIN 30 1183
FT DOMAIN 30 1183
FT TRANSMEM 1158 1177
FT DOMAIN 1178 1183
FT DOMAIN 151 318
FT DOMAIN 533 1093
FT DOMAIN 1093 1157
FT DOMAIN 1151 1156
FT REPEAT 533 719
FT REPEAT 720 906
FT REPEAT 907 1093
FT REPEAT 1151 1156
SQ SEQUENCE 1183 AA; 133066 MW; B6A1CC072E57D76 CRC64;

Query Match
Best Local Similarity 9.7%; Score 151.5; DB 1; Length 1183;
Matches 102; Conservative 45; Mismatches 123; Indels 159; Gaps 27;

QY 2 VENKEETPETPDS-----EETIKANLIFANGSTQTAFFKGTPEKATS 47
Db 713 VTNK-YTPETTSISGEKVDKQDQGRPEKVS--NLLADGEKVKTLT-----VTS 762
QY 48 EA---YAYADTLKKNG-----EYTVVADKGYTLNIFAGKEKTPPEPKEE 91
Db 763 ETNWKYEFKDLPKYDGGKIEYTVTEDHVKDYTTDI--NGTTITNKY-----TPGETSAT 815
QY 92 VT-----IKANLIYADGKTQ-----TAFFKGTPEEATAEAYR 123
Db 816 VTKNDDNNQDGRTEIKVEL-YODGKATGKTALNENNNWTHWTGLDEKAKGQVVK 874
QY 124 Y-ADALKKNGEYTV-DVADKG-----YT-----LNIFAGKEKTPPEPKEEVTKAN 169
Db 875 YTVVELTKVGYTHVDNNDGNLIVTNKYTPETTSISGEKVDKQDQGRPEKVSYN 934
QY 170 LIYADGKTQTAEPKGFEBATAPA---YRYADLLAKENGK-----YTVVDVAD 213
Db 935 LLANGKEKVKTL-----DVTSETNKNYFKDLPKYDEGKIEYTVTEDHVKDYTTDI-- 985
QY 214 KYTLNIFAGKEKTPPEPKEEVT-----IKANLIYADGKTQ----- 250
Db 986 NGTTITNKY-----TPGETSATVTKNWDNNQDGRKRPTEIKVEL-YODGKATGKTALIN 1039
QY 251 -----TAFFKGTFAEATAEAYRY-ADLLAKENGKTYADLED---GGYTTINIRFAGKKVD- 300
Db 1040 ESNWTHWTGLDEKAKGQVVKYTVDELTKVNG-YTHVDNNDGNLIVTNKYTPKPKNK 1098
QY 301 -----EKPEE 305
Db 1099 PIYPERPKD 1107

RESULT 2
EDC8_DAUCA
ID EDC8_DAUCA STANDARD; PRT; 555 AA.
AC P20075;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Embryonic protein DC-8 (Clone 8/10).
GN DC8.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
OX NCBI_TaxID=4039;
```

[illegible]

QY 6 EETPETDSEEEVTKANLIFANGSTQAEFGTSEAYAYADTLK-----KON 60
 Db 1698 EEPSTQNDLSLISV--SQVEASPTSSAHTPSQIASPLQE-----DTLSDVWPDRM 1750
 QY 61 GEYTVADVADKGYTLNKFAGKEKTPPEPKBEVIK--ANLIYADG-----KTQTAEPK 111
 Db 1751 SLVASEKQVSLSE---GEKUSPKSDISPLTPRESSPTSPGFSFSDSGAKESTAAQ 1806
 QY 112 GT-----FEETAAYRYAD-----ALKKNGEYTVADVADKGYT---LNKFAGK 153
 Db 1807 TSSSPPIDAAAEPYGRSSMLFDTMQHLLALSRDLTSSVEKXNGGKTGDFNFYAYQKP 1866
 QY 154 EKTPPEPKE-----EVTKANLI--YADGKTQTAEPFGTSEAYAYADLLAKE 203
 Db 1867 ESTTSPDEEDYDYESHEKTIQADVGVYKTER-----TIKSPDCDSGYSYETI---- 1917
 QY 204 NGKYTVADVADKGYTLNKFAGKEKTPPEPKE-----EVTIKANLIYADGKTQTAEPFGT 258
 Db 1918 -EKTTPPEGGVSCSEI-----TEKTRTPBEGGYSYSEIKT-----TRTPEVSGYT 1964
 QY 259 AEATAEAYRYADLLAKENGKRYTADLEDGYTI 290
 Db 1965 YEKTERSRLLDDIS--NG--YDDTEDGGHTL 1992
 RESULT 7
 MAPB_HUMAN STANDARD; PRT; 2468 AA.
 AC P46821;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain Lc1].
 GN MAP1B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=95104835; PubMed=7806212;
 RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;
 RT "Cloning of human microtubule-associated protein 1B and the identification of a related gene on chromosome 15.";
 RL Genomics 22:273-280(1994).
 CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
 CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN STABILIZING MICROTUBULES.
 CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, Lc1, Lc2 AND Lc3, CAN ASSOCIATE WITH MAP1A AND MAP1B PROTEINS.
 CC -!- DOMAIN: Has a highly basic region with many copies of the sequence KKEE and KKEI/V, repeated but not at fixed intervals, which is responsible for the binding of MAP1B to microtubules.
 CC -!- PTM: Lc1 is coexpressed with MAP1B. It is a polypeptide generated from MAP1B by proteolytic processing. It is free to associate with both MAP1A and MAP1B. It interacts with the amino-terminal region of MAP1B (by similarity).
 CC -!- SIMILARITY: TO MAP1A.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L06237; AAA18904.1; -

DR MIM; 157129; -
 DR InterPro: IPR000102; MAP1B_neuraxin.
 DR Pfam: PF00414; MAP1B_neuraxin; 10.
 DR PROSITE: PS00230; MAP1B_NEURAXIN; 6.
 KW Microtubules; Repeat; Phosphorylation.
 FT CHAIN ? 2468 MAP1 LIGHT CHAIN Lc1.
 FT REPEAT 1878 1894 MAP1B 1.
 FT REPEAT 1895 1911 MAP1B 2.
 FT REPEAT 1912 1928 MAP1B 3.
 FT REPEAT 1929 1945 MAP1B 4.
 FT REPEAT 1946 1962 MAP1B 5.
 FT REPEAT 1963 1979 MAP1B 6.
 FT REPEAT 1997 2013 MAP1B 7.
 FT REPEAT 2014 2030 MAP1B 8.
 FT REPEAT 2031 2047 MAP1B 9.
 FT REPEAT 2048 2064 MAP1B 10.
 FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY KKEE AND KKEI/V REPEATS).
 FT SEQUENCE 2468 AA; 270618 MW; 540839CDBF09D461 CRC64;
 Query Match 8.0%; Score 125; DB 1; Length 2468;
 Best Local Similarity 24.7%; Pred. No. 4.8; Mismatches 104; Indels 72; Gaps 12;
 Matches 71; Conservative 40;
 QY 3 ENKEETPET-----PETDSEEEVTKANLIFANGSTQAEFGTSEAYAYA 53
 Db 560 ESKEETPEVTKVNHVEKPPKVESKEKVMVK-----KDKPVKTETKPSVTEKVPs----- 609
 QY 54 DTLKKONGEYTVADVADKGYTLNKFAGKEK-----PEEPKEE-VTKANLIYADGK 104
 Db 610 ---KEEPSPVKAEVAKQATDVKPKAAKKTVKVKPKEDKKEKPKKEVAKKEDK 666
 QY 105 TQ-TAEFGTTFEATAEAYADALKKNGEYTVADVADKGYTLNKFAGKEKTPPEPKEE 163
 Db 667 TPKKKEPKKEEVKVKV--EIKKEEKKEPKVKV-----KETPPKVEKKE 712
 QY 164 VTIKANLIYADGKTQTAEPFGTSEAYAYADLLAKENGKRYTVADVADKGYTLNKFKA 223
 Db 713 VK-----KEEKKEVKEKKEPKKEIKK---LPKDAKKSSPTLSEAKKPAALK-- 756
 QY 224 GKEKTPPEPKEEVTIKANLIYA-----DGKTQTAEPFGTFAEATAEA 265
 Db 757 -----PKVPKKEESVKKDSVYAGKPKGKIKVIRKEGKAAEAFAAA 798
 RESULT 8
 MAPB_MOUSE STANDARD; PRT; 2464 AA.
 AC P14873;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Microtubule-associated protein 1B (MAP 1B) (MAP1.2) (MAP1(X)) [Contains: MAP1 light chain Lc1].
 GN MAP1B OR MTAP1B OR MTAP5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND DOMAIN.
 RC STRAIN=SWISS WEBSTER; TISSUE=Brain;
 RX MEDLINE=90094539; PubMed=2480963;
 RA Noble M., Lewis S.A., Cowan N.J.;
 RT "The microtubule binding domain of microtubule-associated protein MAP1B contains a repeated sequence motif unrelated to that of MAP2 and tau.";
 RL J. Cell Biol. 109:3367-3376(1989).
 CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
 CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1306;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=G148;
 RX MEDLINE=88029445; PubMed=3665928;
 RA Olsson A., Eliasson M., Guss B., Nilsson B., Hellman U.,
 RA Lindberg M., Uhlen M.;
 RT "Structure and evolution of the repetitive gene encoding
 RT streptococcal protein G.";
 RL Eur. J. Biochem. 168:319-324(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=G148;
 RX MEDLINE=88015586; PubMed=3658689;
 RA Filpula D., Alexander P., Fahnestock S.R.;
 RT "Nucleotide sequence of the protein G gene from Streptococcus GX7805,
 RT and comparison to previously reported sequences.";
 RL Nucleic Acids Res. 15:7210-7210(1987).
 RN [3]
 RP SEQUENCE OF 114-593 FROM N.A.
 RC STRAIN=G148;
 RX MEDLINE=86300657; PubMed=3017704;
 RA Guss B., Eliasson M., Olsson A., Uhlen M., Frej A.-K., Joernvall H.,
 RA Flock J.-I., Lindberg M.;
 RT "Structure of the IgG-binding regions of streptococcal protein G.";
 RL EMBO J. 5:1567-1575(1986).
 RN [4]
 RP STRUCTURE BY NMR OF 371-427.
 RC STRAIN=G148;
 RX MEDLINE=98290449; PubMed=9628485;
 RA Malakauskas S.M., Mayo S.L.;
 RT "Design, structure and stability of a hyperthermophilic protein
 RT variant.";
 RL Nat. Struct. Biol. 5:470-475(1998).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
 CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
 CC IN THE REGION OF THE MEMBRANE ANCHOR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X06173; CAA29540.1; -;
 DR EMBL; Y00428; CAA68489.1; -;
 DR EMBL; X04015; CAA27638.1; -;
 DR PIR; S00128; S00128.
 DR PIR; A26314; A26314.
 DR FDB; IG84; 22-JUL-98.
 DR InterPro; IPR002988; GA.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR000724; IgG_bind_B.
 DR Pfam; PF01468; GA; 3.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR Pfam; PF01378; IgG_binding_B; 3.
 DR PRINTS; PR00015; GPOSANCHOR.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
 KW IgG-binding protein; Repeat; Transmembrane; Cell wall; Signal;
 KW 3D-structure. 1 33
 FT SIGNAL 1 33 IMMUNOGLOBULIN G BINDING PROTEIN G.
 FT CHAIN 34 593 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 34 567 MEMBRANE ANCHOR.
 FT TRANSMEM 568 588
 FT DOMAIN 589 593 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 69 111 ALA-RICH.
 FT DOMAIN 104 290 3 X 37 AA REPEATS.
 FT DOMAIN 104 140 1-1.
 FT REPEAT 104 140 1-1.
 FT REPEAT 179 215 1-2.

FT REPEAT 254 290 1-3.
 FT DOMAIN 303 427 2 X 55 AA REPEATS.
 FT REPEAT 303 357 2-1.
 FT REPEAT 373 427 2-2.
 FT DOMAIN 531 555 5 X 5 AA REPEATS OF [DE]-D-A-K-K.
 FT DOMAIN 559 564 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
 FT PROTEINS.
 SQ SEQUENCE 593 AA; 63291 MW; 048BA760D5B2920 CRC64;
 Query Match 7.8%; Score 122; DB 1; Length 593;
 Best Local Similarity 27.9%; Pred. No. 1.5;
 Matches 72; Conservative 31; Mismatches 87; Indels 68; Gaps 18;
 QY 12 PETDSEEVITKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLK-----KDN---GE 62
 Db 299 PKTDTYK-----LILNGKT---LKG---ETTTEAVDAATAEKVKQYANDNGVDGE 343
 QY 63 YTVVDVADKGYTLNIK---FAGKEKTPPEEPKEVITKANLIYADGKTQTAFFKGTFFFEATA 119
 Db 344 WYDDATKTFVTTEKPEVIDASELTPAVTYKLV-----NGKT---LKG---ETTT 389
 QY 120 EAYRYADALK-----KDN---GEYTVVDVADKGYTLNIK---FAGKEKTPPEEPKEVITK 167
 Db 390 EAVDAATAEKVKQYANDNGVDGEWYDDATKTFVTTEKPEVIDASELTPAVTYKLVIN 449
 QY 168 ANLIYADGKTQTAEFKGTFFFEATAEAYRYADVADLAKEN---GKYTVVDVADKGYTLN---IK 221
 Db 450 GKTL--KGETT--KAVDAETAETAKAFKQ---YANDNGVDGVWYDDATKTFVTTEMVTE 501
 QY 222 FAGKEKTPPEEPKEVITI 238
 Db 502 VPGDAPTEPEKPEASIP 519
 RESULT 11
 ICEV_PSEX
 ID ICEV_PSEX STANDARD; PRT; 1196 AA.
 AC 033479;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein.
 GN INAV.
 OS Pseudomonas syringae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=317;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=INAV5;
 RX MEDLINE=97462815; PubMed=9323042;
 RA Schmid D., Pridmore D., Capitani G., Battistutta R., Neeser J.-R.,
 RA Jann A.;
 RT "Molecular organisation of the ice nucleation protein Inav from
 RT Pseudomonas syringae.";
 RL FEBS Lett. 414:590-594(1997).
 CC -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
 CC CRYSTALLIZATION IN SUPERCOOLED WATER.
 CC -!- SUBUNIT: MEMBRANE ENVIRONMENT OR AGGREGATION SEEMS TO BE REQUIRED
 CC FOR ICE NUCLEATION ACTIVITY.
 CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
 CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
 CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
 CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ001086; CA04521.1; -.
CC HSP: P06620; 11NA.
CC InterPro: IPR000258; Ice_nucleatn.
CC Pfam: PF00818; Ice_nucleatn; 61.
CC PRINTS: PR00327; ICENUCLEATN.
CC PROSITE: PS00314; ICE_NUCLEATION; 42.
CC Ice nucleation; Repeat; Outer membrane.
CC FT DOMAIN 172 1147 OCTAPEPTIDE PERIODICITY.
CC SEQUENCE 1196 AA; 117991 MW; C9E9974CB1731E68 CRC64;

Query Match 7.8%; Score 122; DB 1; Length 1196;
Best Local Similarity 25.0%; Pred. No. 3.3;
Matches 77; Conservative 33; Mismatches 110; Indels 88; Gaps 14;

QY 24 ANLIFANGSTQTAFFK-----GTEKA-----TSEAYAYADTLKDNGEYT 64
Db 712 SSLIAGYGSTQTAGFKSINTAGYGSTQTAQERSDLVAGYGSTSTA-GYSSSLIAGYG--S 768
QY 65 VDVAADKGYTLNFKAGKEKTPPEPKKEVT-----IKANLIYADGKTQTAFFKGTFE 115
Db 769 TQTAGYGSTLTGY-GSTQTAQENSSLTTCYSTAGYSSSLIAGYGSTQTAGYESTL- 826
QY 116 EATAEAYRYADALKKONGEYTVDAKGYTLNFKAGKEKTPPEPKKEVT-----I 166
Db 827 --TA-----GY-----GSTQTAQERSDLVTGYGSTSTAGY 854
QY 167 KANLIYADGKTQTAFFKGTFEATAEAYRYADLLAKENKGYTVDAKGYTLNFKAGKE 226
Db 855 ASLLIAGYGSTQTAGY-----ESTLTAGYGSTQTAQENSSLTGY---GSTAGFASLL 906
QY 227 KTEPEPKKEEYTIKANLIYADGKTQTAFFKGTFAE---ATAEAYRYADLLAKENKGYTADL 283
Db 907 IAGYGSTQTAGYKSTLTLAGYGSTQTAQERSDLTGYGSTATAGQDSSLIAGYGSTLTSGI 966
QY 284 ED---GGY 288
Db 967 RSLTAGY 974

RESULT 12
ID ICEN_PSESY STANDARD; PRT; 1200 AA.
AC P06620;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN INAZ.
OS Pseudomonas syringae (pv. syringae).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas
OX NCBI_TaxID=321;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S203;
RA Green R.L., Warren G.J.;
RT "Physical and functional repetition in a bacterial ice nucleation
RL gene.";
RL Nature 317:645-648(1985).
RN [2]
RP SUBCELLULAR LOCATION.
RX MEDLINE=92297969; PubMed=2520825;
RA Lindow S.E., Lahue E., Govindarajan A.G., Panopoulos N.J., Gies D.;
RT "Localization of ice nucleation activity and the icec gene product in
RL Pseudomonas syringae and Escherichia coli.";
RN [3]
RP Mol. Plant Microbe Interact. 2:262-272(1989).
RP 3D-STRUCTURE MODELING OF 490-535.
RX MEDLINE=93360260; PubMed=8355267;
RA Kajava A.V., Lindow S.E.;
RT "A model of the three-dimensional structure of ice nucleation

```

```

RT proteins.";
RL J. Mol. Biol. 232:709-717(1993).
CC -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC CRYSTALLIZATION IN SUPERCOOLED WATER.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DOMAIN: CONTAINS 122 IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X03035; CAA26837.1; -.
CC PIR: A24405; A24405.
CC PDB: 1INA; 31-OCT-93.
CC PDB: 1INB; 31-OCT-93.
CC InterPro: IPR000258; Ice_nucleatn.
CC Pfam: PF00818; Ice_nucleation; 61.
CC PRINTS: PR00327; ICENUCLEATN.
CC PROSITE: PS00314; ICE_NUCLEATION; 40.
CC Ice nucleation; Repeat; Outer membrane; 3D-structure.
CC FT DOMAIN 176 1151 OCTAPEPTIDE PERIODICITY.
CC SEQUENCE 1200 AA; 118587 MW; E3BEAD4BEA67ABB0 CRC64;

Query Match 7.8%; Score 122; DB 1; Length 1200;
Best Local Similarity 24.2%; Pred. No. 3.3;
Matches 74; Conservative 33; Mismatches 115; Indels 84; Gaps 12;

QY 24 ANLIFANGSTQTAFFK-----TFEKATSEAYAYADTLKKNGEYTVDAKGYT 73
Db 716 SSLIAGYGSTQTAGFKSILTAGYGSTQTAQERTSLVAGYGST--STAGYSSSLIAGYGST 773
QY 74 LNIKFA-----GKEKTPPEPKKEVT-----IKANLIYADGKTQTAFFKGTFEA 117
Db 774 QTAGYESTLTAGYGSTQTAQENSSLTTCYSTAGYSSSLIAGYGSTQTAGYESTL--- 830
QY 118 TAEAYRYADALKKONGEYTVDAKGYTLNFKAGKEKTPPEPKKEVT-----IKA 168
Db 831 TA-----GY-----GSTQTAQERSDLVTGYGSTSTAGYAS 860
QY 169 NLIYADGKTQTAFFKGTFEATAEAYRYADLLAKENKGYTVDAKGYTLNFKAGKEK 228
Db 861 SLIAGYGSTQTAGY-----ESTLTAGYGSTQTAQENSSLTGY---GSTAGFASLLIS 912
QY 229 PEEPKEEYTIKANLIYADGKTQTAFFKGTFAE---ATAEAYRYADLLAKENKGYTADLE 285
Db 913 GYGSTQTAGYKSTLTAGYGSTQTAQENSSLTTCYSTAGYGSTATAGQDSSLIAGYGS 972
QY 286 ---GGY 288
Db 973 FLTAGY 978

RESULT 13
SSP5_STRGN
ID SSP5_STRGN STANDARD; PRT; 1500 AA.
AC P16952; Q54184;
DT 01-AUG-1990 (Rel. 15, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Agglutinin receptor precursor (SSP-5).
GN SSP5 OR SSPB.
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

```

Streptococcus.
NCBI_TaxID=1302;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=M5;
RX MEDLINE=90236997; PubMed=2185241;
RA Demuth D.R., Golub E.E., Malamud D.;
RT "Streptococcal-host interactions. Structural and functional analysis
of a Streptococcus sanguis receptor for a human salivary
glycoprotein.";
RL J. Biol. Chem. 265:7120-7126(1990).
[2]
RN REVISIONS, SEQUENCE FROM N.A.
RC STRAIN=M5;
RX MEDLINE=963310377; PubMed=8733238;
RA Demuth D.R., Duan Y., Brooks W., Holmes A.R., McNab R.,
RJ Jenkinson H.F.;
RT "Random genes encode cell-surface polypeptides SspA and SspB which
mediate adhesion of the oral bacterium Streptococcus gordonii to
human and bacterial receptors.";
RL Mol. Microbiol. 20:403-413(1996).
-1- FUNCTION: MAY BIND SALIC ACID.
CC (SAG) IN A CALCIUM-DEPENDENT REACTION. THE INTERACTION OF SAG WITH
CC ITS RECEPTOR IN VARIOUS ORAL STREPTOCOCCI MODULATE BACTERIAL
CC COLONIZATION OF ORAL TISSUE AND IS ASSOCIATED WITH REDUCED LEVELS
CC OF DENTAL CARIES.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -1- DOMAIN: THE PR2 REGION, BY SIMILARITY WITH THE PROLINE RICH
CC DOMAINS OF THE S.PYOGENES M6 PROTEIN AND STAPHYLOCOCCAL PROTEIN A,
CC MAY TRAVERSE THE CELL WALL PEPTIDOGLYCAN AND IS FOLLOWED BY
CC HYDROPHOBIC AA WHICH MAY FUNCTION TO ANCHOR THE PROTEIN TO THE
CC MEMBRANE.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -1- SIMILARITY: BELONGS TO THE SPAP/SSP5/SPAA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U40026; AAC44100.1; --
CC PIR; A35186; A35186.
CC InterPro: IPR001899; Gram_pos_anchor.
CC Pfam; PF00746; Gram_pos_anchor; 1.
CC PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
CC Signal; Repeat; Calcium-binding; Transmembrane.
CC SIGNAL 1 38
CC CHAIN 39 1500 AGGLUTININ RECEPTOR.
CC DOMAIN 39 1474 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 1475 1495 POTENTIAL.
CC DOMAIN 1496 1500 CYTOPLASMIC (POTENTIAL).
CC STIMILAR 164 470 TO M PROTEIN OF S.PYOGENES.
CC DOMAIN 164 470 4 X APPROXIMATE TANDEM REPEATS, HR1.
CC REPEAT 164 241 1.
CC REPEAT 242 323 2.
CC REPEAT 324 405 3.
CC REPEAT 406 470 4.
CC DOMAIN 771 887 PRO-RICH (PR2).
CC DOMAIN 1414 1436 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
CC DOMAIN 1467 1472 PROTEINS.
CC CA_BIND 220 235 POTENTIAL.
CC CA_BIND 301 316 POTENTIAL.
CC CA_BIND 931 950 POTENTIAL.
CC CA_BIND 1300 1315 POTENTIAL.
CC SEQUENCE 1500 AA; DCF190E7D4D889F CRC64;
Query Match 7.8%; Score 121.5; DB 1; Length 1500;
Best Local Similarity 25.3%; Pred. No. 4.5;

Matches 76; Conservative 35; Mismatches 134; Indels 55; Gaps 13;
QY 3 ENKEETPETDSEEEVITKANLIFANGSTQTAEEFGTPEKATSEAYAYADTLKKDNGE 62
DB 114 ETDKDGTATTATDQAQ-----KQDEIKSDYAKQAEIEITTEAYKKEVAHQOATDKINAE 169
QY 63 YTVADVADKGYTLNKFAGKERTPEEPKEEVTIKANLIYADCKTQTAEFGTFFEEATAEAY 122
DB 170 N--KAADKKYQKDL-----KSHQEEVEKI-----NTANATAKAEYEAQLAQYQ 210
QY 123 RYADALAKDNGEYTVADVADKGYTLNKFAGKERTPEEPKE--EVTIKANLIYADG----- 175
DB 211 KDLATVKKANEDSQDYQNKLSAYQTELARVQKANAKEAYEKAVKENTAKNAKLVEN 270
QY 176 --KTQTAEEFGTFFEEATAEAYRVADLLA--KENGKTYTVADVADKGYTLNKFAGKERTPE 230
DB 271 EATKQNETAKATY--EAMKQYE--ADLAAIKKANEDNDADYQAKLAAYQTELARVQKANA 328
QY 231 EPKE--EVTIKANLIYADGKTQTAEEFGTPEAEATAEAYRVADLLAKEN-----GKYTADL 283
DB 329 EAKEAYDKAVKEN-----TA--KNTAIQAEANEAKORNETAKATYDAAYKKYEADL 377
RESULT 14
MRSP_STAAU STANDARD; PRT; 1637 AA.
ID MRSP_STAAU STANDARD; PRT; 1637 AA.
AC P80544; Q92F62;
DT 01-FEB-1996 (Rel. 33, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Methicillin-resistant surface protein precursor.
GN PLS.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=ISOLATE 1061;
RA Savolainen K., Kuusela P., Paulin L., Korhonen T.K.;
RT "Pls, a large repeat-rich surface protein of methicillin resistant
RT Staphylococcus aureus.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 624-628; 676-682; 938-948; 1156-1168; 1176-1185;
RP 1199-1205 AND 1217-1224.
RC STRAIN=ISOLATE 1061;
RX MEDLINE=96270743; PubMed=8665912;
RA Hilden P., Savolainen K., Tynnelae J., Vuontola M., Kuusela P.;
RT "Purification and characterisation of a plasmin-sensitive surface
RT protein of Staphylococcus aureus.";
RL Eur. J. Biochem. 236:904-910(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF115379; AAD09131.1; --
CC InterPro: IPR001899; Gram_pos_anchor.
CC Pfam; PF00746; Gram_pos_anchor; 1.
CC PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
CC Cell wall; Transmembrane; Antibiotic resistance; Glycoprotein; Repeat;
CC Signal.
CC SIGNAL 1 48 POTENTIAL.
CC CHAIN 49 1637 METHICILLIN-RESISTANT SURFACE PROTEIN.
CC DOMAIN 1301 1582 141 X 2 AA TANDEM REPEATS OF D-[SAG].
CC DOMAIN 1598 1603 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
CC SEQUENCE 1637 AA; 174573 MW; 75BE9ADB469BD309 CRC64;

Query Match 7.8%; Score 121.5; DB 1; Length 1637;
Best Local Similarity 27.3%; Pred. No. 4.9;
Matches 79; Conservative 20; Mismatches 111; Indels 79; Gaps 12;

QY 3 ENKEETPETDSEEEVTIKANLIIFANGSTQTAEFKGTPEKATSEAYAYADTLKKDNGE 62
DB EQVDVTKDTTEQASTE--KAN-----TTEQASTEAKADTTEQATTEAPKAEG- 133
QY 63 YTVDVADKGYTLNKFAGKEKTPPEEKKEVTIKANLIYADGKTQTAFFKGTFFEATAEAY 122
DB 134 -----TDKVTETEAAPAEETDKAT-----TEEPKAEETDKATEAP 170
QY 123 RYADALKDNGEYTVDVADKGYTLNKF--GKEKTPPEEKKEVTIKANLIYA-----D 174
DB 171 K-----TEETDKATTEAPAEETSAAATEAPKAETSAAATEAPKAET 217
QY 175 GKTQTAFFKGTFFEE---ATAEAYRYADLLAKENGKTVTVDVADKGYTLNKFAGKEKTPPE 231
DB 218 EKTAATEAPKTEETDKVTEAPK-----AEETSAAATEAPKAETN-----KVETEEA 267
QY 232 PKEEVTIKANLIYADGKTQTAFFKGTFAEATA-----EAYRYADLLAKE 275
DB 268 PAAEETNKA-----ATEETPAVEDTNAKSNSNAOPSETERTQVVYDVVAKD 312

RESULT 15
ABPX_YEAST STANDARD; PRT; 627 AA.
AC Q08641; Q08644;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Actin-binding protein ABP140.
GN ABP140 OR YOR239W/YOR240W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=97127829; PubMed=8972580;
RA Boyer J., Michaux G., Fairhead C., Gaillon L., Dujon B.;
RT "Sequence and analysis of a 26.9 kb fragment from chromosome XV of
the yeast Saccharomyces cerevisiae.";
RL Yeast 12:1575-1586(1996).
RN [2]
RP PARTIAL SEQUENCE FROM N.A., SEQUENCE OF 1-12; 38-72; 102-115; 244-286;
RP 560-574 AND 598-608, FUNCTION, AND SUBCELLULAR LOCATION.
RC STRAIN=B35457;
RX MEDLINE=98127445; PubMed=9467951;
RA Asakura T., Sasaki T., Nagano F., Satch A., Obaishi H., Nishioaka H.,
RA Imamura H., Hotta K., Tanaka K., Nakanishi H., Takai Y.;
RT "Isolation and characterization of a novel actin filament-binding
protein from Saccharomyces cerevisiae.";
RL Oncogene 16:121-130(1998).
CC -!- FUNCTION: Binds F-actin and shows weak F-actin crosslinking
activity.
CC -!- SUBCELLULAR LOCATION: Cytoplasm and cortical cytoskeleton.
CC -!- MISCELLANEOUS: A RIBOSOMAL FRAMESHIFT OCCURS BETWEEN THE CODONS
FOR LEU-276 AND GLY-277.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 275147; CAA99460.1; ALT_SEQ.
DR EMBL; 275147; CAA99461.1; ALT_SEQ.

DR SGD; S0005765; ABP140.
DR InterPro; IPR001601; Meth-transf.
DR InterPro; IPR000051; SAM_bind.
KW Actin-binding; Cytoskeleton; Ribosomal frameshift.
FT INIT_MET 0
SQ SEQUENCE 627 AA; 71354 MW; D4E55F9485412F39 CRC64;

Query Match 7.7%; Score 121; DB 1; Length 627;
Best Local Similarity 22.5%; Pred. No. 1.9;
Matches 64; Conservative 52; Mismatches 103; Indels 66; Gaps 12;

QY 4 NKEETPETDSEEEVTIKANLIIFANGSTQTAEFKGTPEKATSEAYAYADTLKK-DNGE 62
DB 13 SKEEGDATVDTNSSK-PLKS-----DETKELHQESTAVPQEVDMVEEFENEP 61
QY 63 YTVD---VADKGYTLNKFAGKEKTPPEEKKEVTIKANLIYADGKT-----QTAEF 110
DB 62 ETINSRTAEKPLETNLP---KPEETNEEDEFEGSMSENKIYSGENADINVNDFOEYKEM 118
QY 111 KGTFFEATAEAYRYADALKKDNCEYTVDVADKGYTLNKFAGKEKTPPEEKKEVTIKANL 170
DB 119 ENTGAELASSVEESDAIQEGVAETEGIA-----TPKOKENEKNDESEE 163
QY 171 IYADGKTQTA-EFKGTFFEATAEAYRYADLLAKENGKTYVDV-----ADKGYT----L 218
DB 164 ESANNASEPAAEYSQSEADADIE-----QSNKETENAEASQAQANDGSTSTTSK 214
QY 219 NIKFAGKEKTPPEEKKEVTIKANLIYADGKTQTAEFKGTFAEATA 263
DB 215 NKKKKNNKKNNKRRNGNVNTNANV---DDSTKTGENDDTTGDTTTS 256

Search completed: October 29, 2002, 09:29:54
Job time : 14.429 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:27:36 ; Search time 12.7943 Seconds
(without alignments)
2290.646 Million cell updates/sec

Title: US-08-325-278B-1

Perfect score: 1565

Sequence: 1 AVENKEETPETDSEEV.....GGYTINIRFAGKKVDEKPEE 305

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1565	100.0	455	2 A45063	immunoglobulin-bin
2	1561	99.7	719	2 A42808	Ig light chain-bin
3	1326.5	78.4	932	2 S54396	protein L precursor
4	171.5	11.0	74	2 A34483	Ig light chain-bin
5	167	10.7	664	2 S42574	streptococcal surf
6	166	10.6	479	2 T47561	late embryogenesis
7	151.5	9.7	1185	2 A42404	collagen adhesin -
8	151	9.6	448	2 H84782	late embryogenesis
9	149	9.5	448	2 JC6171	late embryogenesis
10	141.5	9.0	1110	2 I51116	NF-180 - sea lamp
11	141	9.0	555	2 S04909	embryonic protein
12	137.5	8.8	992	2 E72350	hypothetical prote
13	132.5	8.5	737	2 H81070	lactoferrin-bindin
14	131.5	8.4	891	2 G41662	130K surface exclu
15	131.5	8.4	1166	2 T28680	fibrinogen-binding
16	130.5	8.3	701	2 H98120	choline binding pr
17	130.5	8.3	1141	2 E98824	hypothetical prote
18	130.5	8.3	1337	2 T30291	dextranase - Strep
19	128.5	8.2	938	2 AF1772	internalin-like pr
20	127.5	8.1	843	2 S72375	surface exclusion
21	127.5	8.1	1940	2 A29320	myosin heavy chain
22	127	8.1	863	2 S06017	neuraxin - rat
23	127	8.1	913	2 T52485	neurofilament prot
24	127	8.1	2364	2 A56577	microtubule-associ
25	126.5	8.1	1035	2 T16588	hypothetical prote
26	126	8.1	955	2 A60990	alpha-amylase (EC
27	125.5	8.0	673	2 C86278	F14L17.15 protein
28	125.5	8.0	1526	2 A45605	mature-parasite-in
29	125	8.0	463	2 S27757	embryonic abundant

30 124.5 8.0 1302 1 JC6009 surface-located me
31 124 7.9 495 2 D89808 hypothetical prote
32 124 7.9 1489 2 T31108 cyst germination s
33 123.5 7.9 1198 2 T42223 ladder protein - C
34 123.5 7.9 2464 1 QRMSPI microtubule-associ
35 123 7.9 569 2 T39577 hypothetical prote
36 123 7.9 1822 2 S33441 EF protein - Strep
37 122.5 7.8 698 2 S52696 myosin heavy chain
38 122.5 7.8 890 2 S22452 surface exclusion
39 122.5 7.8 955 2 S24348 myosin heavy chain
40 122.5 7.8 1993 2 AF1450 probable peptidogl
41 122 7.8 332 2 C40646 endospore developm
42 122 7.8 384 2 G86287 hypothetical prote
43 122 7.8 593 2 S00128 protein G precurs
44 122 7.8 1200 1 SNPSO ice nucleation pro
45 121.5 7.8 657 2 AD1525 probable cell surf

ALIGNMENTS

RESULT 1

A45063

immunoglobulin-binding protein LG - Peptostreptococcus magnus

C;Species: Peptostreptococcus magnus

C;Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999

C;Accession: A45063

R;Kihlberg, B.M.; Sjobring, U.; Kastern, W.; Bjorck, L.

J. Biol. Chem. 267, 25583-25588, 1992

A;Title: Protein LG: a hybrid molecule with unique immunoglobulin binding properties.

A;Reference number: A45063; MUID:93094283

A;Accession: A45063

A;Status: preliminary

A;Molecule type: mRNA; protein

A;Residues: 1-455 <Kih>

A;Cross-references: GB:S50809; NID:g261705; PID:AAA03280.1; PID:g261706

A;Note: sequence extracted from NCBI backbone (NCBIN:120302, NCBIP:120303)

Query Match 100.0%; Score 1565; DB 2; Length 455;

Best Local Similarity 100.0%; Pred. No. 2e-85;

Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	AVENKEETPETDSEEV	TIKANLIFANGSTQTA	EFKGTAEKATSEAYADTLKKN	60
Db	22	AVENKEETPETDSEEV	TIKANLIFANGSTQTA	EFKGTAEKATSEAYADTLKKN	81
Qy	61	GEYTVADVADKGYTLN	IKFAGKEKTPPEPK	EEVTIKANLIYADGKTQTA	120
Db	82	GEYTVADVADKGYTLN	IKFAGKEKTPPEPK	EEVTIKANLIYADGKTQTA	141
Qy	121	AYRYADALKDNGEYTV	DVADKGYTLNKPAGKE	KTPEEPKEEVTIRKANLIYADGKTQTA	180
Db	142	AYRYADALKDNGEYTV	DVADKGYTLNKPAGKE	KTPEEPKEEVTIRKANLIYADGKTQTA	201
Qy	181	EFGKTFEATAEAYRY	ADLLAKENGYTVADV	ADKGYTLNPKFAGKEKTPPEPK	240
Db	202	EFGKTFEATAEAYRY	ADLLAKENGYTVADV	ADKGYTLNPKFAGKEKTPPEPK	261
Qy	241	NLIYADGKTQTAEPKG	TFEATAEAYRYADLL	AKENGYTADLEDGGYTTINIRFAGK	300
Db	262	NLIYADGKTQTAEPKG	TFEATAEAYRYADLL	AKENGYTADLEDGGYTTINIRFAGK	321
Qy	301	EKPEE	305		
Db	322	EKPEE	326		

RESULT 2

A42808

Ig light chain-binding protein precursor - Peptostreptococcus magnus

N;Alternate names: protein L

C;Species: Peptostreptococcus magnus

C;Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 07-May-1999

DB 241 MERKLEKSEKTEPE--PEEEVITKANLIFADGSTQNAEFKGTFAKAVSDAYAYADALKKDNG 298

Db 214 DAPELTPALTYKLIVKGN--TFSGETTTKAITATAAEKFKQ-YATANNV---DGWESYD 268

QY 67 VADKGYTLNKAFA--GKEKTPPEKPEEVIKANLIYADGKTOTAEFGTPEEATAEAYR 123

Db 269 DAYKTFVTTEKPAVIDAPELTPALTYKLVKGNTP--SGETTT---KAVDAETAERAKFK 323

QY 124 -YADALKKONGEYTDVADKGYTLINKAFA--GKEKTPPEKPEEVIKANLIYADGKTOT 179

Db 324 QYATANNVD--GWSYDDATKTFVTTEKPAVIDAPELTPALTYKLVKGNTP--SGETTT 380

QY 180 AEFKGTPEEATAEAYRYADLLAKENGKTYTDVADKGYTLINKAFA--GKEKTPPEKPEE 236

Db 381 KAIDAATAEAEKFKQYATANGV---DGWESYDDATKTFVTTEKPAVIDAPELTPALTYKL 437

QY 237 TIKANLIYADGKTOTAEFGTPEEATAEAYRYADLLAKENGKTYTDADGCGYTINIRFAG 296

Db 438 IVKGNTP--SGETTT---KAVDAETAERAKFKQ---YANENGVI-----GEWSYDDATKT 483

QY 297 KKVDEKP 303

Db 484 FTVTEKP 490

RESULT 6

T47561

late embryogenesis abundant protein-like - Arabidopsis thaliana

N:Alternate names: protein F8J2.210

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T47561

R:Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.; Mayer, K.F.X.

submitted to the Protein Sequence Database, April 2000

A:Reference number: Z24458

A:Accession: T47561

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-479 <N>A>

A:Cross-references: EMBL:AL132969

A:Experimental source: cultivar Columbia; BAC clone F8J2

C:Genetics:

A:Map position: 3

A:Introns: 329/3; 382/2

A:Note: F8J2.210

Query Match 10.6%; Score 166; DB 2; Length 479;

Best Local Similarity 25.8%; Pred. No. 0.0072;

Matches 82; Conservative 43; Mismatches 137; Indels 56; Gaps 12;

QY 3 ENKEETPETPTDSEEEVITKANLIFANGSTQTAEFKGTPEKATSEAYAY---ADFLKK 58

Db 75 DTAESTREGADIASEKAAGR-----DTTGEVRDSTAQKTETADYADKAREAKD 125

QY 59 DNGEYTVDVADKGYTLINIKFAGKEKTPPEKPEEVIKANLIYADGKTOTAEFGTPEEAT 118

Db 126 KTADKTRETAD--YAAEKAREAKDRTADTKETAETAYAEKA-REAKDKTADKLGEYKDYT 182

QY 119 ABAYRYA-DALKDONGEYTDVADKGYTLINIKFAGKEKTPPEKPEEVIKANLIYADGKT 177

Db 183 AEKAKEAKDTTAEKLGEY-----KDYTVDKAKEAKDKTAEKATAEYTSDKA-RETKD 235

QY 178 QTAEFKGTPEEATAE-----AYRYADLLAKENGKTYTDVADKGYTLINIKFAGKEKTP 229

Db 236 KTAKEKVEYKDYTAEKAKETADKAREAKDKTAEKVGRIYDYTAEK---ATTCKDGVSKIG 293

QY 230 EEPKEEV-TIKANLIYADGKTQ-----TAEFGTGFABATAEAYRYADLLAKENGKTYADL 283

Db 294 ELKDSAVDTAKRANGFLSGKTEETKQKAVETKDTAKEKMDGEAGEARRKKEE----- 345

QY 284 EDGCGYTINIRFAGKVKDE 301

Db 346 -----MRLEGGKKLDE 355

RESULT 7

A42404

collagen adhesin - Staphylococcus aureus

C:Species: Staphylococcus aureus

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 10-Nov-1995

C:Accession: A42404; S27665

R:Patti, J.M.; Jonsson, H.; Guss, B.; Switalski, L.M.; Wiberg, K.; Lindberg, M.; Hook J.; Biol. Chem. 267, 4766-4772, 1992

A:Title: Molecular characterization and expression of a gene encoding a Staphylococcus A:Reference number: A42404; MUID:92165839

A:Contents: FDA 574

A:Accession: A42404

A:Molecule type: DNA

A>Status: preliminary

A:Residues: 1-1185 <PAT>

A:Cross-references: EMBL:M81736

A:Note: sequence extracted from NCBI backbone (NCBIP:83982)

Query Match 9.7%; Score 151.5; DB 2; Length 1185;

Best Local Similarity 23.8%; Pred. No. 0.15;

Matches 102; Conservative 45; Mismatches 123; Indels 159; Gaps 27;

QY 2 VENKEETPETPTDSE-----EEVTIRKANLIFANGSTQTAEFKGTPEEKATS 47

Db 715 VTNK-VTPETTSISGEKVDKDNQDKRPEKYSV--NLLADGEKVKTLD-----VTS 764

QY 48 EA---YAYADTLKKDNG-----EYTVDVADKGYTLINIKFAGKEKTPPEEK 91

Db 765 ETNWKYEFKDLPRYDEGKIEYTVTEHDVADYTTDI--NGTTITNKY-----TPGETSAT 817

QY 92 VT-----IKANLIYADGKTQ-----TAEFGTPEEATAEAYR 123

Db 818 VTNNDDNNQDKRPTTEIKVEL-YDQKATGKTALINESNNWTHWTGLDEKAKGQVK 876

QY 124 Y-ADALKKONGEYTDVADKG-----YT-----LNIRFAGKEKTPPEEVIKAN 169

Db 877 YTVEELTKVKGYTHVDNDMGNLIVTNKYPTPETTSISGEKVDKDNQDKRPEKYSVN 936

QY 170 LIYADGKTQTAEPKGFEEATAEA---YRYADLLAKENGK-----YTVDVAD 213

Db 937 LLANGEKVKTL-----DVTSETNWKYEFKDLPRYDEGKIEYTVTEHDVADYTTDI-- 987

QY 214 KGYTLINIKFAGKEKTPPEEKPEEYV-----IKANLIYADGKTQ----- 250

Db 988 NGTITNKY-----TPGETSATVTKNWDDNNQDKRPTTEIKVEL-YDQKATGKTALIN 1041

QY 251 -----TAEFGTGFABATAEAYRY-ADLLAKENGKTYADLED---GGYTINIRFAGKVKVD- 300

Db 1042 ESNWTHWTGLDEKAKGQVKYTVDELTKVNG-YTTHVDNDMGNLIVTNKYTPKKPNK 1100

QY 301 -----EKPEE 305

Db 1101 PIYPEKPKD 1109

RESULT 8

H84782

late embryogenesis abundant protein (AtECP63) [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: H84782

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y. M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, Natus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: H84782

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-448 <STO>

A:Cross-references: GB:AE002093; NID:g4415909; PIDN:AAD20140.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g36640

A:Map position: 2

Query Match 9.6%; Score 151; DB 2; Length 448;

Best Local Similarity 24.5%; Pred. No. 0.051;

Matches 79; Conservative 38; Mismatches 107; Indels 98; Gaps 14;

QY 3 ENKEETPETDSEEVITKANLIANGSTQTAFFKGTFAEATAYADTLKKDNGE 62

DB 79 EAAESTKEGAQIASEKAVGAK-----DAIVKAKET-----ADYTAEKVGE 119

QY 63 YTVADVADKGYTLNIFAGKEKTPPEKPEEVIKANLIYADGKTQTAFFKGTFAEATAEAY 122

DB 120 Y-----KDYTVDKAKEAKDTTAEKAKETANYTADKA-VEAKDKTAEKIGEYK----- 165

QY 123 RYADALKKNGEYTVADVADKGYTLNIFAGKEKTPPEKPEEVIKANLIYADGKTQTAFF 182

DB 166 -----DYAVDKA-----VEAKDKTAEKAKETANYTADKA-KEAKDKTAEK 204

QY 183 KGTFEEATAE-----AYRVADLLAKENGKGYTVADVADKGYTLNIFAGKEKTPPE 230

DB 205 VGEYKDYTVDKAVEADYDTAEKAEIAEKDKTAEKIGEY-----KDYTVKATEGKDVTVS 258

QY 231 ---EPKEEV---TIKANLIYADGKTQTAFFKGTFAEATAEAYRVADLLAKENGKGYTADLED 285

DB 259 KLGLKDSAVETAKRAMGFLSGKTEEAAGKAVETKDT-----AKEN-----MEK 302

QY 286 GGYTI-----NIRFAGKKVDEK 302

DB 303 AGEVTRQKMEEMRLGKELKEE 324

RESULT 9

JC6171

late embryogenesis abundant protein - Arabidopsis thaliana

N:Alternate names: embryogenic cell protein 63; phosphotyrosine protein

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 21-Jul-2000

A:Accession: JC6171

R:Yang, H.; Saitou, T.; Kameda, Y.; Harada, H.; Kamada, H.

Gene 184, 83-88, 1997

A:Title: Arabidopsis thaliana ECP63 encoding a LEA protein is located in chromosome 4.

A:Reference number: JC6171; MUID: 97169149

A:Accession: JC6171

A:Molecule type: mRNA

A:Residues: 1-448 <YAN>

A:CROSS-references: DDBJ:D641140; NID:gl526423; PIDN:BAAL1017.1; PID:gl526424

C:Comment: This protein is a phosphotyrosine protein which belongs to late embryogenesis

C:Genetics:

A:Gene: ATECP63

A:Map position: 4, south part

C:Keywords: seed

Query Match 9.5%; Score 149; DB 2; Length 448;

Best Local Similarity 24.5%; Pred. No. 0.067;

Matches 79; Conservative 38; Mismatches 107; Indels 98; Gaps 14;

QY 3 ENKEETPETDSEEVITKANLIANGSTQTAFFKGTFAEATAYADTLKKDNGE 62

DB 79 EAAESTKEGAQIASEKAVGAK-----DAIVKAKET-----ADYTAEKVGE 119

QY 63 YTVADVADKGYTLNIFAGKEKTPPEKPEEVIKANLIYADGKTQTAFFKGTFAEATAEAY 122

DB 120 Y-----KDYTVDKAKEAKDTTAEKAKETANYTADKA-VEAKDKTAEKIGEYK----- 165

QY 123 RYADALKKNGEYTVADVADKGYTLNIFAGKEKTPPEKPEEVIKANLIYADGKTQTAFF 182

DB 166 -----DYAVDKA-----VEAKDKTAEKAKETANYTADKA-KEAKDKTAEK 204

QY 183 KGTFEEATAE-----AYRVADLLAKENGKGYTVADVADKGYTLNIFAGKEKTPPE 230

DB 205 VGEYKDYTVDKAVEADYDTAEKAEIAEKDKTAEKIGEY-----KDYTVKATEGKDVTVS 258

QY 231 ---EPKEEV---TIKANLIYADGKTQTAFFKGTFAEATAYADTLKKDNGE 285

DB 259 KLGLKDSAVETAKRAMGFLSGKTEEAAGKAVETKDT-----AKEN-----MEK 302

QY 286 GGYTI-----NIRFAGKKVDEK 302

DB 303 AGEVTRQKMEEMRLGKELKEE 324

RESULT 10

IS1116

NF-180 - sea lamprey

C:Species: Petromyzon marinus (sea lamprey)

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C:Accession: IS1116

R:Jacobs, A.J.; Kamholz, J.; Selzer, M.E.

Brain Res. Mol. Brain Res. 29, 43-52, 1995

A:Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation

A:Reference number: IS1116; MUID: 95287814

A:Accession: IS1116

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1110 <JAC>

A:CROSS-references: EMBL:U19361; NID:g632548; PIDN:AAA80106.1; PID:g632549

C:Superfamily: neurofilament triplet H protein

Query Match 9.0%; Score 141.5; DB 2; Length 1110;

Best Local Similarity 25.4%; Pred. No. 0.54;

Matches 87; Conservative 38; Mismatches 138; Indels 79; Gaps 13;

QY 1 AVENKEETPETDSEEVITKANLIANGSTQTAFFKGTPE-KATSEAYAYADTLKKD 59

DB 482 SAEKDEEEEEEKEEEE-----AAEEEEEEDGRKEGEAAEEAEVEEKEE 532

QY 60 NGEYTVADVADKGYT-----LNIFAGKEKTPPEKPEEVIKANLIYADGKT 105

DB 533 AEEAEVEEAETEAEEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 584

QY 106 QTAFFKGTFAEATAE-----YRVADLLKNGEYTVADVADKGYTLNIFAGKEKT 156

DB 585 EAAEAKAEVEEEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 638

QY 157 PEE--PKEEVIKANLIYADGKTQTAFFKGTFAEATAYRVADLLAKENGKGYTVADVADK 214

DB 639 EEEAEAEAEVTSK-----KAKTQAE-----VEEEAEAEAEAEAEAEAEAE 684

QY 215 GYTLNIFAGKEKTPPEPK-----EVTIKANLIYADGKTQTAFFKGTFAEATAE 264

DB 685 -----VEAESEKEEEDSKDAEAEAEAEAEAEAEAEAEAEAEAEAEAE 739

QY 265 AYRYA-DLLAKENGKGYTADLEDGGYTNIRFAGKKVDEKPEE 305

DB 740 AAEEAKDEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 781

RESULT 11

S04909

embryonic protein DC8 (clone 8/10) - carrot

C:Species: Daucus carota (carrot)

C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 21-Jul-2000

C:Accession: S04909

R:Franz, G.; Hatzopoulos, P.; Jones, T.J.; Krauss, M.; Sung, Z.R.

Mol. Gen. Genet. 218, 143-151, 1989

A:Title: Molecular and genetic analysis of an embryonic gene, DC 8, from Daucus carot

A:Reference number: S04909; MUID: 89384429

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-555 <FRA>

A:CROSS-references: GB:X16131; NID:gl18333; PIDN:CAA34258.2; PID:g4902464

Query Match

9.0%; Score 141; DB 2; Length 555;

Best Local Similarity 25.2%; Pred. No. 0.26;
Matches 78; Conservative 30; Mismatches 130; Indels 72; Gaps 11;
QY 6 EETPETPETSBEVITKANLIFANGSTQTAEPKGTAEKATSEAYAYA-----DTLK 57
Db 182 QKAAEAKDTTAAK-----AEAKETGYKYAAKAAEAKVLAQAQAAEAKDITG 232
QY 58 KNGEYTVADVADKGYTLNIFAGKKTPEEPKEEVTIKANLIYADGKTQ-TAEFKGTFFEE 116
Db 233 KD-GEYKYAAQK--AAEAKDATMQKTGE-----YKDYAAQKTAETKDATWE 276
QY 117 ATAEAYRYA-----DALKKNGEYTVADVADKGYTLNIFAGKKTPEEPKEEVTIKA 168
Db 277 KAKEYREYAAKAAEAKDATMQKTGEY-----KDYSAQKAAETKDATMEKTE----- 324
QY 169 NLIYADGKTQTA-----EFKGTFFEEATAEAYRYADLLAKENGKYYTVADVADKGY 216
Db 325 ---YKDYTAQAAETKDATMEKAKEAKDTTVQKTGEYKYDYAAEKAGKGVTVVEKAKEG 381
QY 217 TLNIFAGKKTPEEPKEEVTIKANLIYADGKTQTAEPKGTAEKATSEAYRYADLLAKEN 276
Db 382 DTTV---GKMTLKDAAARKAMDMLGKKEEVKAGETAEEAAKEKYEDTEFAARKK 438
QY 277 GKYTADLEDG 286
Db 439 MEELKLEEG 448

RESULT 12
E72350
hypotheical protein TM0638 - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
R:Nelson, K.B.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316
A:Accession: E72350
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-992 <ARN>
A:Cross-references: GB:AE001738; GB:AE000512; NID:g4981158; PIDN:AAD35722.1; PID:g498116
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0638

Query Match 8.8%; Score 137.5; DB 2; Length 992;
Best Local Similarity 22.3%; Pred. No. 0.81;
Matches 77; Conservative 40; Mismatches 112; Indels 117; Gaps 15;
QY 2 VENKEETPETPDSBEVITKANLIFANGSTQTAEE-----FKGTPEKATYS 47
Db 219 VLSKDETVEYSEKDIYOK-----NLLKPGDTVHVQKIDERFAYVVGAVARPGMYTFSRE 273
QY 48 EAYAYADTLKDKNG-----EYTVADVADKGYTLNI-----KFAG 80
Db 274 ESUTLNLIAGAGVSVDKRFIEKALITRGGETLEVDPPVLDENVEIEGVDVVEIKKF-- 331
QY 81 KEKTEP-----EPKEEVTIKANLIYADGKTQTAEPKGTFFEEATAEAYRYAD 126
Db 332 -EXTEVVSQVSRPGVYEISPKENTYLEKLLSMAGS-----IKGTLEVD-----SI 378
QY 127 ALKKNGEYTVADVADKGYTLNIFAGKKTPEEPKEEVTIKANLIYADGKTQTAEPKGTFF 186
Db 379 VVTRDGSVITLS-----PNNLDFSVKPGDVNVNVEFVPKKA---YVLGVVRN---PGLY 426
QY 187 EATAEAYRYADLLAKENGKYYTVADVADKGYTLNIFAGKKTPEEPKEEVTIKANLIYAD 246
Db 427 TFGKNEAFTLRNLIAGAG-----FVDEQGVSVKVGAKYEYSPDE-----IVKEDILL 476

QY 247 GKTQTAEPKGTFAEATAEAYRYADLL-----AKENGKRYTADLED 285
Db 477 GYFVVYE-----RYTDRFVYVWGVNVRNGKMSPEKEE 509
RESULT 13
H81070
lactoferrin-binding protein B NMB1541 [imported] - Neisseria meningitidis (strain MC5
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: H81070
R:Tetzelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755
A:Accession: H81070
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-737 <TET>
A:Cross-references: GB:AE002504; GB:AE002098; NID:g7226785; PIDN:AAF41896.1; PID:g722
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1541

Query Match 8.5%; Score 132.5; DB 2; Length 737;
Best Local Similarity 22.8%; Pred. No. 1.1;
Matches 85; Conservative 50; Mismatches 137; Indels 101; Gaps 17;
QY 5 KEETPETPET-DSEEEVTIK-ANLIFANGSTQTAEPKGTAEKATSEAYA--YADTLKKDN 60
Db 81 KQDGTIPDKHQAEEHLPLKXDILFDGLTK--EQADKLKKINERYSDVRVITSKKEE 138
QY 61 GEYTVADVADKGYTLNI---KFAKKEKTPPEEPKEEVTIKA--NLIYADGK-----TQT 107
Db 139 EKYQYQFVRAGYVYFTRAEGKDNEKEKT-SDGKEFVNRFSYDGVYVYSGERPSQSLPSAGT 197
QY 108 AEPKGTFE-----EATAEAYRYADALLKDKNGEYTVD 138
Db 198 VOYSGNWQYMTDAKRHTGKAVSSDGLGYTYTYGNEIGATSYEARDADREKHPAEYTVD 257
QY 139 VADKGYTLNIK-----FAGKEKTPPEEPKEEVTIKANLIYADGKTQTAEPKGTPEEATAEA 193
Db 258 FDNK--TLNGKLKNQYVQKSNPNPKRPLTIYDITATLDGNRFTGSAKVSTEVKTQHA 315
QY 194 ---YRYADLLAKENGKYYTVADVADKGYTLNIK-----FAGKEKTPPEEPKEE- 236
Db 316 DREYLFHTDADQRLGEGF---FGDNGEELAGRFISNDNSVFGVFGAKQKTETENAADTK 372
QY 237 -----TIKANLIYADGKTQTAEPKGTFAEATAEAYRYADLLAKE-----N 276
Db 373 PALSSGKHVKILDSLKISVDEASDKNPR-----EFAISSMPDPGHPDKLLVEGREIPLVN 427
QY 277 GKYTADLEDGGYT 289
Db 428 KEQTIELADGRKT 440

RESULT 14
G41662
130K surface exclusion protein Sec10 precursor - Enterococcus faecalis plasmid pCF10
C:Species: Enterococcus faecalis
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 15-Oct-1999
C:Accession: G41662
R:Kao, S.M.; Olmsted, S.B.; Viksnius, A.S.; Gallo, J.C.; Dunny, G.M.
J. Bacteriol. 173, 7650-7664, 1991
A:Title: Molecular and genetic analysis of a region of plasmid pCF10 containing posit
terococcus faecalis.
A:Reference number: A41662; MUID:92041679
A:Accession: G41662
A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-891 <RAO>

A:Cross-references: GB:M64978; NID:g150552; PIDN:AAA65847.1; PID:g150554

C:Genetics:

A:Genome: plasmid

Query Match

Best Local Similarity 8.4%; Score 131.5; DB 2; Length 891;

Matches 78; Conservative 46; Mismatches 148; Indels 81; Gaps 11;

```
QY 5 KETPETPDSSEETIKANLIFANGSTQTAFFKGTPEKATSEAYAYADTLKKNGEY 64
   :: ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 29 EQAPKTPENSTEQPTVKTQTTEQAITEKQQQVTEKQAIYDQKQVADTAKKDAID 88
   :: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 65 VDVAAGYTLNIFAGKEKTPPEKPEEVIKANLIYADGKTQTAFFKGTPEEATAEAY 124
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 89 QSVKDDQAVVD-----QNKDALDSQAVT-----DQAVVDEAKKVDEATPSAIEK 136
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 125 A-----DALKKDNGEYTVDVADGKGYTLNIFAGKEKTPPEKPEEVIKANLIYADGKTQ 178
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 137 AKQVATDQAVDEQKQVVDQAGTD-----VNOQAVVDEKAKETNAKV--QNEKDQ 187
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 179 TAEFKGTPEEATAEAYRYADLLAKENGKTVTVADGKGYTLNIFAGKEKTPPEKPEEVI 238
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 188 QAVTAAKQEQAKLEE-----LAKNAEAEKVAEKEQAAKEAEELANKQKEAKAKDKT- 240
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 239 KANLIYADGKT--QTAFFKGTFAEATAEAYRYADLLAKENG----- 277
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 KDDQAVADQGTVTVTTSQEKVTDKADATA--KQADLTAKENALKDKQAATKQAOQNTLDNSK 299
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 278 -----KYTAD-----LEDGGYTIINIRFAGKKVDEKPEE 305
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 300 EELKGHGINLPKFSADYDTKLSEAEIATLEKTALEMKNKNEPTSKDEKNKD 352
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 15

T28680

fibrinogen-binding protein homolog - Staphylococcus aureus

C:Species: Staphylococcus aureus

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-May-2000

C:Accession: T28680

R:Josefsson, E.; McCrea, K.; Ni Eidhin, D.; O'Connell, D.; Cox, J.; Hook, M.; Foster, T.

Microbiology 144, 3387-3395, 1998

A:title: Three new members of the serine-aspartate repeat protein multigene family of Staphylococcus aureus

A:Reference number: Z20510; MUID:99098700

A:Accession: T28680

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1166 <JOS>

A:Cross-references: EMBL:AJ005647; NID:el318793; PID:el318794; PIDN:CAA06652.1

C:Genetics:

A:Gene: sdre

Query Match

Best Local Similarity 8.4%; Score 131.5; DB 2; Length 1166;

Matches 82; Conservative 46; Mismatches 134; Indels 139; Gaps 17;

```
QY 13 ETDSEEEVITKANLIFANGSTQTAFFKGTPEKATSEAYAYADTLKKDNGEYTVDAV- 68
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 414 ETSQNTVDYQDPMVHGDSNIQSIFTKLDEDKQTIQQIYVNPPLKKSATNTKVDIAGSQV 473
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 69 -DKGYTLNIFAGKEKTPPEKPEEVI-----KANLIY----- 100
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 474 DDYG---NIKL-GNGSTIIDQNTKIKVKNVNSDQQLPQSNRIYDFSQYEDVTSQFDNKK 529
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 101 -----ADGKTQTAFFKGTPEEATAEAYRYADALKKN 132
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 530 FSNVATLDFGINSAYIIKWSKYTPTSDGELDIAQ--GT-SMRTTDKYGYNYAGYSN 586
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 133 -----GEYTVDVADGKGYTL-----NIKFAKKEKTPPEKPEEVIKANLIYADGK 176
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 587 FIYTSNDTGGGDTGVRPEEKLYIGDYVVEDVDKGVQGDTSKSRPMANVLTLTYPDGT 646
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:28:27 ; Search time 9.49256 Seconds
(without alignments)
784.805 Million cell updates/sec

Title: US-08-325-278B-1

Perfect score: 1565

Sequence: 1 AVENKETPTPTDSEEV.....GGYTINIRFAGKKVDEKPEE 305

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2.6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2.6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2.6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2.6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2.6/ptodata/1/iaa/PTUS_COMB.pep.*
- 6: /cgn2.6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1565	100.0	305	2	US-08-795-475-1
2	1565	100.0	434	2	US-08-795-475-3
3	1226.5	78.4	1027	4	US-08-446-137B-2
4	1216	77.7	291	4	US-08-446-137B-4
5	618	39.5	342	3	US-08-828-741B-6
6	618	39.5	342	4	US-09-160-567-6
7	615	39.3	178	3	US-08-828-741B-13
8	615	39.3	178	4	US-09-160-567-13
9	615	39.3	198	3	US-08-828-741B-8
10	615	39.3	198	4	US-09-160-567-8
11	614	39.2	495	3	US-08-828-741B-4
12	614	39.2	495	4	US-09-160-567-4
13	613	39.2	182	3	US-08-828-741B-2
14	613	39.2	182	4	US-09-160-567-2
15	334	21.3	75	4	US-08-446-137B-8
16	315.5	20.2	71	4	US-08-446-137B-6
17	306	19.6	74	4	US-08-446-137B-7
18	303	19.4	71	4	US-08-446-137B-5
19	167	10.7	664	3	US-08-669-408B-2
20	151.5	9.7	1183	2	US-08-447-031A-2
21	128.5	8.2	564	4	US-09-308-022-6
22	127.5	8.1	631	4	US-08-847-065-25
23	123	7.9	936	5	PCT-US93-05944-2
24	121.5	7.8	522	4	US-08-961-083-120
25	121.5	7.8	1040	4	US-08-961-083-118
26	116.5	7.4	1912	1	US-08-409-995-4
27	116.5	7.4	1912	3	US-08-685-467-4

28	116.5	7.4	2353	4	US-09-377-155-33	Sequence 33, Appl
29	116.5	7.4	2353	4	US-08-913-942-4	Sequence 4, Appli
30	116.5	7.4	2353	4	US-09-669-974-33	Sequence 33, Appl
31	116.5	7.4	2354	4	US-09-268-347-47	Sequence 47, Appl
32	116	7.4	1104	4	US-09-268-347-28	Sequence 28, Appl
33	116	7.4	1104	4	US-09-268-347-34	Sequence 34, Appl
34	115	7.3	272	3	US-08-441-857-4	Sequence 4, Appli
35	115	7.3	272	3	US-08-193-159-4	Sequence 4, Appli
36	115	7.3	283	4	US-09-367-012-1	Sequence 1, Appli
37	115	7.3	337	2	US-08-441-857-8	Sequence 8, Appli
38	115	7.3	337	3	US-08-193-159-8	Sequence 8, Appli
39	115	7.3	365	3	US-08-483-577A-156	Sequence 156, App
40	115	7.3	365	4	US-08-897-438-156	Sequence 156, App
41	115	7.3	404	3	US-08-483-577A-155	Sequence 155, App
42	115	7.3	404	4	US-08-897-438-155	Sequence 155, App
43	115	7.3	411	3	US-08-483-577A-154	Sequence 154, App
44	115	7.3	411	4	US-08-897-438-154	Sequence 154, App
45	115	7.3	417	3	US-08-483-577A-153	Sequence 153, App

ALIGNMENTS

RESULT 1
US-08-795-475-1
; Sequence 1, Application US/08795475
; Patent No. 5965390
; GENERAL INFORMATION:
; APPLICANT: Bjvrck, Lars
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,475
; FILING DATE: 11-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mcmasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.402D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 305 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli LE392/PHDL, DSM 7054
US-08-795-475-1

DNAdiomed
seq2

Query Match 100.0%; Score 1565; DB 2; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.6e-126;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AVENKETPTPTDSEEVITKANLIFANGSTQTAFFKCTFEKATSEAYAYADTLKKDN 60
|||||

Db 1 AVENKETPTETDSEEEVTIKANLIFANGSTQTAFFKGTFFKATSEAYAYADTLKKDN 60

QY 61 GEYTVADVADKGYTLNKFAGKEKTPPEPKEEVTIKANLIYADGKTQTAFFKGTFFKATSEAYAYADTLKKDN 120

Db 61 GEYTVADVADKGYTLNKFAGKEKTPPEPKEEVTIKANLIYADGKTQTAFFKGTFFKATSEAYAYADTLKKDN 120

QY 121 AYRYADALKKNDGEYTVADVADKGYTLNKFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 180

Db 121 AYRYADALKKNDGEYTVADVADKGYTLNKFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 180

QY 181 EFKGTFFKATSEAYAYADLAKENGYTVADVADKGYTLNKFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 240

Db 181 EFKGTFFKATSEAYAYADLAKENGYTVADVADKGYTLNKFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 240

QY 241 NLIYADGKTQTAFFKGTFFKATSEAYAYADLAKENGYTVADVADKGYTLNKFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 300

Db 241 NLIYADGKTQTAFFKGTFFKATSEAYAYADLAKENGYTVADVADKGYTLNKFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 300

QY 301 EKPEE 305

Db 301 EKPEE 305

RESULT 2

US-08-795-475-3

; Sequence 3, Application US/08795475

; Patent No. 5965390

; GENERAL INFORMATION:

; APPLICANT: Bjvrck, Lars

; APPLICANT: Sjrving, Ulf

; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED AND BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/795,475

; FILING DATE: 11-FEB-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: McMasters, David D.

; REGISTRATION NUMBER: 33,963

; REFERENCE/DOCKET NUMBER: 100084.402D1

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 434 amino acids

; TYPE: amino acid

; STRANDEDNESS: unknown

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ORIGINAL SOURCE:

; ORGANISM: Escherichia coli LE392/pHDLG, DSM 7055

US-08-795-475-3

Query Match 100.0%; Score 1565; DB 2; Length 434;

Best Local Similarity 100.0%; Pred. No. 2.7e-126;

Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVENKETPTETDSEEEVTIKANLIFANGSTQTAFFKGTFFKATSEAYAYADTLKKDN 60

Db 1 AVENKETPTETDSEEEVTIKANLIFANGSTQTAFFKGTFFKATSEAYAYADTLKKDN 60

QY 61 GEYTVADVADKGYTLNKFAGKEKTPPEPKEEVTIKANLIYADGKTQTAFFKGTFFKATSEAYAYADTLKKDN 120

Db 61 GEYTVADVADKGYTLNKFAGKEKTPPEPKEEVTIKANLIYADGKTQTAFFKGTFFKATSEAYAYADTLKKDN 120

QY 121 AYRYADALKKNDGEYTVADVADKGYTLNKFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 180

Db 121 AYRYADALKKNDGEYTVADVADKGYTLNKFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 180

QY 181 EFKGTFFKATSEAYAYADLAKENGYTVADVADKGYTLNKFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 240

Db 181 EFKGTFFKATSEAYAYADLAKENGYTVADVADKGYTLNKFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 240

QY 241 NLIYADGKTQTAFFKGTFFKATSEAYAYADLAKENGYTVADVADKGYTLNKFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 300

Db 241 NLIYADGKTQTAFFKGTFFKATSEAYAYADLAKENGYTVADVADKGYTLNKFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 300

QY 301 EKPEE 305

Db 301 EKPEE 305

RESULT 3

US-08-446-137B-2

; Sequence 2, Application US/08446137B

; Patent No. 6162903

; GENERAL INFORMATION:

; APPLICANT: Trowern, Angus R.

; APPLICANT: Atkinson, Anthony

; APPLICANT: Murphy, Jonathan P.

; APPLICANT: Laurence, Oliver S.

; APPLICANT: Duggleby, Clive J.

; TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED

; TITLE OF INVENTION: FROM L PROTEIN AND THEIR USES

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED AND BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/446,137B

; FILING DATE: 22-MAY-1995

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: McMasters, David D.

; REGISTRATION NUMBER: 33,963

; REFERENCE/DOCKET NUMBER: 100084.406

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1027 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-446-137B-2

Query Match 78.4%; Score 1226.5; DB 4; Length 1027;

Best Local Similarity 79.9%; Pred. No. 7.7e-97;

Matches 246; Conservative 22; Mismatches 31; Indels 9; Gaps 4;

QY 2 VENKETPTETDSEEEVTIKANLIFANGSTQTAFFKGTFFKATSEAYAYADTLKKONG 61

Db	2	ETPE--PEEEVTIKANLIFADGSGTQNAEPKGTFAKAVSDAYAYADALKKDNGEYTVDVAD	59
Qy	70	KGYTLNFKAGKEBTPPEPKEEVTIKANLIYADGKTQTAEFGTFPEATAEAYRYADALX	129
Db	60	KGLTLNFKAGKKEPPEPKEEVIKNVLIFADGKTQTAEFGTFPEATAKAYADLLA	119
Qy	130	KDNGEYTVDVADKGTYLNLIKAFGE--KTPPEPKPEEVTIKANLIYADGKTQTAEFGTFFE	187
		:	
Db	120	KENSEYTADLEDGGNTINIKFAGKETPETPEEPKEEVTIKVNLFADGKIQTAEFGTFFE	179
Qy	188	EATAEAYRYADLLAKENGYTVDOVGKYTLNIKFKAGE--KTPPEBPKEEVTIKANLIA	245
Db	180	EATAKAYANLLAKENGEYTADLEDGGNTINIKFAGKETPETPEEPKEEVTIKVNLI	239
Qy	246	DGKTQTAEFGTFFAEATAEAYRYADLLAKENGYKTADLEDGGYTNIRFAGK	297
Db	240	DGKTQTAEFGTFFAEATAEAYRYADLLAKVENGYTTADLEDGGYTNIRKIFAGK	291

RESULT 5
US-08-828-741B-6

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

```

? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Digiglio, Frank S.
? REGISTRATION NUMBER: 31, 346
? REFERENCE/DOCKET NUMBER: 10591
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (516) 742-4343
? TELEFAX: (516) 742-4366
? TELEX: 230 901 SANS UR
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 342 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-828-741B-6

Query Match 39.5%; Score 618; DB 3; Length 342;
Best Local Similarity 75.9%; Pred. No. 1.9e-45;
Matches 126; Conservative 5; Mismatches 15; Indels 20; Gaps

Qy 16 SEEEVTKANLIFANGSTQTAEEKGTFEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 75
|
Db 174 SAEVTKANLIFANGSTQTAEEKGTFEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 233
|

```

[illegible]

```

Db 234 IKFAGKEATNRNTDGDSTDYGILOINSRWGLTSAEEVTIKANLIFANGSTQTAEFKGTPE 293
QY ~ 116 EATAEAYADALKKNGEYTVDVADKGYTLNKFAGKEKTPPEPK 161
      :||:|||| ||| ||||||||| ||||||||| ||| :|
Db 294 KATSEAYAYADTLKKDNGEYTVDVADKGYTLNKFAGKESAWRHPQ 339

RESULT 6
US-09-160-567-6
: Sequence 6, Application US/09160567
: Patent No. 6326179
: GENERAL INFORMATION:
: APPLICANT: Koentgen, Frank
: APPLICANT: Suess, Gabriele M.
: APPLICANT: Tarlinton, David M.
: APPLICANT: Treutlein, Herbert R.
: TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
: STREET: 400 Garden City Plaza
: CITY: Garden City
: STATE: New York
: COUNTRY: United States of America
: ZIP: 11530
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/160,567
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/828,741
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Digiglio, Frank S.
: REGISTRATION NUMBER: 31,346
: REFERENCE/DOCKET NUMBER: 10591
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (516) 742-4343
: TELEFAX: (516) 742-4366
: TELEX: 230 901 SANS UR
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 342 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-160-567-6

Query Match 39.5%; Score 618; DB 4; Length 342;
Best Local Similarity 75.9%; Pred. No. 1.9e-45;
Matches 126; Conservative 5; Mismatches 15; Indels 20; Gaps 1;

QY 16 SEEEVTIKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 75
      | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 174 SAEVTIKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 233
      :||:|||| ||| ||||||||| ||||||||| ||| :|
QY 76 IKFAGKEKTPPEE-----PKEEVTIKANLIYADGKTQTAEFKGTPE 115
      ||||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 234 IKFAGKEATNRNTDGDSTDYGILOINSRWGLTSAEEVTIKANLIFANGSTQTAEFKGTPE 293
QY 116 EATAEAYADALKKNGEYTVDVADKGYTLNKFAGKEKTPPEPK 161
      :||:|||| ||| ||||||||| ||||||||| ||| :|
Db 294 KATSEAYAYADTLKKDNGEYTVDVADKGYTLNKFAGKESAWRHPQ 339

RESULT 7
US-08-828-741B-13

```

```

: Sequence 13, Application US/08828741B
: Patent No. 6043069
: GENERAL INFORMATION:
: APPLICANT: Koentgen, Frank
: APPLICANT: Suess, Gabriele M.
: APPLICANT: Tarlinton, David M.
: APPLICANT: Treutlein, Herbert R.
: TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
: STREET: 400 Garden City Plaza
: CITY: Garden City
: STATE: New York
: COUNTRY: United States of America
: ZIP: 11530
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/828,741B
: FILING DATE: 26-MAR-1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Digiglio, Frank S.
: REGISTRATION NUMBER: 31,346
: REFERENCE/DOCKET NUMBER: 10591
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (516) 742-4343
: TELEFAX: (516) 742-4366
: TELEX: 230 901 SANS UR
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 178 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-828-741B-13

Query Match 39.3%; Score 615; DB 3; Length 178;
Best Local Similarity 76.2%; Pred. No. 1.4e-45;
Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1;

QY 18 EEVTIKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 77
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12 EEVTIKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 71
      :||:|||| ||| ||||||||| ||||||||| ||| :|
QY 78 FAGKEKTPPE-----PKEEVTIKANLIYADGKTQTAEFKGTPEEA 117
      ||||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 72 FAGKEATNRNTDGDSTDYGILOINSRWGLTSAEEVTIKANLIFANGSTQTAEFKGTPEKA 131
      :||:|||| ||| ||||||||| ||||||||| ||| :|
QY 118 TAEAYADALKKNGEYTVDVADKGYTLNKFAGKEKTPPEPK 161
      :||:|||| ||| ||||||||| ||||||||| ||| :|
Db 132 TSEAYAYADTLKKDNGEYTVDVADKGYTLNKFAGKESAWRHPQ 175

RESULT 8
US-09-160-567-13
: Sequence 13, Application US/09160567
: Patent No. 6326179
: GENERAL INFORMATION:
: APPLICANT: Koentgen, Frank
: APPLICANT: Suess, Gabriele M.
: APPLICANT: Tarlinton, David M.
: APPLICANT: Treutlein, Herbert R.
: TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER

```

STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: United States of America
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/160,567
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/828,741
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 178 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-160-567-13

Query Match 39.3%; Score 615; DB 4; Length 178;
Best Local Similarity 76.2%; Pred. No. 1.4e-45;
Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1;

QY 18 EEVTKANLIFANGSTQTAEFGKTFEKATSEAYAYADTLKKDNGEYTVADVADKGYTLNIK 77
Db 12 EEVTKANLIFANGSTQTAEFGKTFEKATSEAYAYADTLKKDNGEYTVADVADKGYTLNIK 71
QY 78 FAGKEKTPPE-----PKEEVTKANLIYADGKTOTAEFGKTFEEA 117
Db 72 FAGKEATNRNTDGTGYILQINRWGLTSAEEVTKANLIFANGSTQTAEFGKTFEKA 131
QY 118 TAAAYRYADALKKDNGEYTVADVADKGYTLNIKAGKEKTPPEPK 161
Db 132 TSEAYAYADTLKKDNGEYTVADVADKGYTLNIKAGKESAWRHPQ 175

RESULT 9
US-08-828-741B-8
Sequence 8, Application US/08828741B
Patent No. 6043069
GENERAL INFORMATION:
APPLICANT: Koentgen, Frank
APPLICANT: Suess, Gabriele M.
APPLICANT: Tarlinton, David M.
APPLICANT: Treutlein, Herbert R.
TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
OPERATING SYSTEM: PRODUCING SAME
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: United States of America
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/828,741B
FILING DATE: 26-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 198 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-828-741B-8

Query Match 39.3%; Score 615; DB 3; Length 198;
Best Local Similarity 76.2%; Pred. No. 1.6e-45;
Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1;

QY 18 EEVTKANLIFANGSTQTAEFGKTFEKATSEAYAYADTLKKDNGEYTVADVADKGYTLNIK 77
Db 32 EEVTKANLIFANGSTQTAEFGKTFEKATSEAYAYADTLKKDNGEYTVADVADKGYTLNIK 91
QY 78 FAGKEKTPPE-----PKEEVTKANLIYADGKTOTAEFGKTFEEA 117
Db 92 FAGKEATNRNTDGTGYILQINRWGLTSAEEVTKANLIFANGSTQTAEFGKTFEKA 151
QY 118 TAAAYRYADALKKDNGEYTVADVADKGYTLNIKAGKEKTPPEPK 161
Db 152 TSEAYAYADTLKKDNGEYTVADVADKGYTLNIKAGKESAWRHPQ 195

RESULT 10
US-09-160-567-8
Sequence 8, Application US/09160567
Patent No. 6326179
GENERAL INFORMATION:
APPLICANT: Koentgen, Frank
APPLICANT: Suess, Gabriele M.
APPLICANT: Tarlinton, David M.
APPLICANT: Treutlein, Herbert R.
TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
OPERATING SYSTEM: PRODUCING SAME
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: United States of America
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/160,567
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/828,741
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10591

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 198 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-160-567-8

Query Match      39.3%; Score 615; DB 4; Length 198;
Best Local Similarity 76.2%; Pred. No. 1.6e-45;
Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1;

QY 18 EEEVTKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 77
DB 32 EEEVTKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 91
QY 78 FACKGKTPPE-----PKEEVTKANLIYADGKTOTAEFKGTFEEA 117
DB 92 FACKGKTPPE-----PKEEVTKANLIYADGKTOTAEFKGTFEEA 151
QY 118 TAPAYRYADALKKDNGEYTVDVADKGYTLNFKAGKTPPEPK 161
DB 152 TSEAYAYADTLKKDNGEYTVDVADKGYTLNFKAGKTPPEPK 195

RESULT 11
US-08-828-741B-4
; Sequence 4, Application US/08828741B
; Patent No. 6043069
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Suess, Gabriele M.
; APPLICANT: Tarlington, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; TITLE OF INVENTION: PRODUCING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 26-MAR-1997
; APPLICATION NUMBER: US/08/828,741B
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 495 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-828-741B-4

Query Match      39.3%; Score 615; DB 4; Length 198;
Best Local Similarity 76.2%; Pred. No. 1.6e-45;
Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1;

QY 18 EEEVTKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 77
DB 32 EEEVTKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 91
QY 78 FACKGKTPPE-----PKEEVTKANLIYADGKTOTAEFKGTFEEA 117
DB 92 FACKGKTPPE-----PKEEVTKANLIYADGKTOTAEFKGTFEEA 151
QY 118 TAPAYRYADALKKDNGEYTVDVADKGYTLNFKAGKTPPEPK 161
DB 152 TSEAYAYADTLKKDNGEYTVDVADKGYTLNFKAGKTPPEPK 195

RESULT 12
US-09-160-567-4
; Sequence 4, Application US/09160567
; Patent No. 6326179
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Suess, Gabriele M.
; APPLICANT: Tarlington, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; TITLE OF INVENTION: PRODUCING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/09/160,567
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/828,741
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 495 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-160-567-4

Query Match      39.2%; Score 614; DB 4; Length 495;
Best Local Similarity 78.6%; Pred. No. 6.8e-45;
Matches 125; Conservative 4; Mismatches 10; Indels 20; Gaps 1;

QY 16 SEEEVTKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 75
DB 174 SAEVTKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 233
QY 76 IKFAGKTPPE-----PKEEVTKANLIYADGKTOTAEFKGTFFE 115
DB 234 IKFAGKTPPE-----PKEEVTKANLIYADGKTOTAEFKGTFFE 293
QY 116 EATAPAYRYADALKKDNGEYTVDVADKGYTLNFKAGRE 154
DB 294 KATSEAYAYADTLKKDNGEYTVDVADKGYTLNFKAGRE 332

Query Match      39.2%; Score 614; DB 4; Length 495;
Best Local Similarity 78.6%; Pred. No. 6.8e-45;
Matches 125; Conservative 4; Mismatches 10; Indels 20; Gaps 1;

QY 16 SEEEVTKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 75
DB 174 SAEVTKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 233
QY 76 IKFAGKTPPE-----PKEEVTKANLIYADGKTOTAEFKGTFFE 115

```

```

||||| I
Db 234 IKFAGKEATNRNTDGDYDYLQINRWGGLTSAEEVTIKANLIFANGSTQTAEFKGTFF 293
QY 116 EATAEAYRYADALKKNGEYTVDVADKGYTLINIKFAGKE 154
Db 294 KATSEAYAYADTLKKNGEYTVDVADKGYTLINIKFAGKE 332

RESULT 13
US-08-828-741B-2
; Sequence 2, Application US/08828741B
; Patent No. 6043069
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Suess, Gabriele M.
; APPLICANT: Tarlinton, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; TITLE OF INVENTION: PRODUCING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828, 741B
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELEPHONE: (516) 742-4366
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 182 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-828-741B-2

Query Match 39.2%; Score 613; DB 3; Length 182;
Best Local Similarity 78.5%; Pred. No. 2.le-45;
Matches 124; Conservative 5; Mismatches 9; Indels 20; Gaps 1;

QY 15 DSEEVTIKANLIFANGSTQTAEFKGTSEAYAYADTLKKNGEYTVDVADKGYTL 74
Db 25 DNTBEVTIKANLIFANGSTQTAEFKGTSEAYAYADTLKKNGEYTVDVADKGYTL 84

QY 75 NIKFAGKEKTPPEP-----KEEVTIKANLIYADGKTQTAEFKGTFF 114
Db 85 NIKFAGKEATNRNTDGDYDYLQINRWGGLTLKEEVTIKANLIFANGSTQTAEFKGTFF 144

QY 115 EATAEAYRYADALKKNGEYTVDVADKGYTLINIKFAG 152
Db 145 EKATSEAYAYADTLKKNGEYTVDVADKGYTLINIKFAG 182

RESULT 14
US-09-160-567-2
; Sequence 2, Application US/09160567
; Patent No. 6326179

```

```

; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Suess, Gabriele M.
; APPLICANT: Tarlinton, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; TITLE OF INVENTION: PRODUCING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/160,567
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/828,741
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 182 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-160-567-2

Query Match 39.2%; Score 613; DB 4; Length 182;
Best Local Similarity 78.5%; Pred. No. 2.le-45;
Matches 124; Conservative 5; Mismatches 9; Indels 20; Gaps 1;

QY 15 DSEEVTIKANLIFANGSTQTAEFKGTSEAYAYADTLKKNGEYTVDVADKGYTL 74
Db 25 DNTBEVTIKANLIFANGSTQTAEFKGTSEAYAYADTLKKNGEYTVDVADKGYTL 84

QY 75 NIKFAGKEKTPPEP-----KEEVTIKANLIYADGKTQTAEFKGTFF 114
Db 85 NIKFAGKEATNRNTDGDYDYLQINRWGGLTLKEEVTIKANLIFANGSTQTAEFKGTFF 144

QY 115 EATAEAYRYADALKKNGEYTVDVADKGYTLINIKFAG 152
Db 145 EKATSEAYAYADTLKKNGEYTVDVADKGYTLINIKFAG 182

RESULT 15
US-08-446-137B-8
; Sequence 8, Application US/08446137B
; Patent No. 6162903
; GENERAL INFORMATION:
; APPLICANT: Trowern, Angus R.
; APPLICANT: Atkinson, Anthony
; APPLICANT: Murphy, Jonathan P.
; APPLICANT: Laurence, Oliver S.
; APPLICANT: Duggleby, Clive J.
; TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED
; TITLE OF INVENTION: FROM L PROTEIN AND THEIR USES
; NUMBER OF SEQUENCES: 12

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,137B
FILING DATE: 22-MAY-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: McMasters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 100084.406
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 75 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-446-137B-8

Query Match 21.3%; Score 334; DB 4; Length 75;
Best Local Similarity 90.1%; Pred. No. 4.5e-22;
Matches 64; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 227 KTEPEKPEVTIKANLIYADGKTQTAEFGKGTFAEATAEAYRYADLLAKENGKYTABLEDG 286
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 5 ETPPEKPEVTIKVNLIFADGKTQTAEFGKGTPEATAEAYRYADLLAKVNGEYTABLEDG 64

QY 287 GYTINIRFAGK 297
|||||:|||||
Db 65 GYTINIRFAGK 75
```

Search completed: October 29, 2002, 09:31:54
Job time : 11.4926 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:25:06 ; Search time 23.9378 Seconds
(without alignments)
1415.233 Million cell updates/sec

Title: US-08-325-278B-1
Perfect score: 1565
Sequence: 1 AVENKEETPETDSEEV.....GGYTINIRFAGKKVDEKPE 305

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1565	100.0	305	14	AA42993
2	1565	100.0	434	14	AA42994
3	1544.5	98.7	467	22	AA31372
4	1366	87.3	367	21	AA10432
5	1226.5	78.4	1027	14	AA42203
6	1226.5	78.4	1027	14	AA43699
7	1216	77.7	291	14	AA42204
8	618	39.5	342	18	AA32482
9	618	39.5	342	20	AA406909
10	615	39.3	178	18	AA42486
11	615	39.3	178	20	AA406913

1999 date
- 2000 date

12	615	39.3	198	18	AAW32481	Growth factor LHL.
13	615	39.3	198	20	AAW06910	Amino acid sequenc
14	614	39.2	495	18	AAW32480	Growth factor CATA
15	614	39.2	495	20	AAW06908	CATAB-TEV aminoaci
16	613	39.2	182	18	AAW32479	Growth factor LHL
17	613	39.2	182	20	AAW06907	LHL growth factor
18	608	38.8	482	20	AAW06915	Amino acid sequenc
19	389	24.9	76	21	AAW82537	Peptostreptococcus
20	371	23.7	72	21	AAW82538	Peptostreptococcus
21	370	23.6	72	21	AAW82540	Peptostreptococcus
22	369	23.6	72	21	AAW82539	Peptostreptococcus
23	341	21.8	82	21	AAW82536	Ig light chain bin
24	334	21.3	75	21	AAW82544	Peptostreptococcus
25	334	21.3	82	21	AAW82545	PpL mutant protein
26	334	21.3	82	21	AAW82548	PpL mutant protein
27	332	21.2	82	21	AAW82546	PpL mutant protein
28	328	21.0	82	21	AAW82547	PpL mutant protein
29	315.5	20.2	71	21	AAW82542	Peptostreptococcus
30	306	19.6	74	21	AAW82543	Peptostreptococcus
31	303	19.4	71	21	AAW82541	Peptostreptococcus
32	167	10.7	664	16	AAW71929	S. dysgalactiae MI
33	154.5	9.9	1185	13	AAW22675	Collagen binding p
34	149	9.5	448	21	AAW77965	A. thaliana enviro
35	141.5	9.0	1638	20	AAW00138	Enterococcus faeca
36	141.5	9.0	1638	20	AAW00140	Enterococcus faeca
37	141.5	9.0	1638	20	AAW00142	Enterococcus faeca
38	131.5	8.4	414	20	AAW49150	Amino acid sequenc
39	131.5	8.4	1166	20	AAW08643	S. aureus SdrE pro
40	130.5	8.3	376	20	AAW49231	N-terminal truncat
41	130.5	8.3	376	20	AAW32103	Choline binding pr
42	130.5	8.3	413	20	AAW49247	N-terminal region
43	130.5	8.3	413	20	AAW32186	N-terminal choline
44	130.5	8.3	414	20	AAW49144	Amino acid sequenc
45	130.5	8.3	564	19	AAW62654	C3 binding protein

ALIGNMENTS

RESULT 1						
AA42993	ID	AA42993	standard; Protein; 305 AA.			
XX	AC	AAW2993;				
XX	XX	16-MAY-1994	(first entry)			
XX	XX	Immunoglobulin light chain binding protein (Protein L).				
DE	DE	Immunoglobulin; light chain; binding; identification; purification;				
KW	KW	separation.				
XX	OS	E. coli LE392/pHDL, DSM 7054.				
XX	XX	Key	Location/Qualifiers			
FT	FT	Domain	5..305			
FT	FT		/label= B1 immunoglobulin light chain binding			
FT	FT	Domain	81..305			
FT	FT		/label= B2 immunoglobulin light chain binding			
FT	FT	Domain	153..305			
FT	FT		/label= B3 immunoglobulin light chain binding			
FT	FT	Domain	225..305			
FT	FT		/label= B4 immunoglobulin light chain binding			
FT	FT	Domain	297..305			
FT	FT		/label= B5 immunoglobulin light chain binding			
XX	XX	WO9322342-A.				

```

PD 11-NOV-1993.
XX
PF 28-APR-1993; 93WO-SE00375.
XX
PR 28-APR-1992; 92SE-0001331.
XX
PA (HIGH-) HIGHTECH RECEPTOR AB.
XX
PI Bjoerck L, Sjoerbring U;
XX WPI; 1993-368722/46.
XX DR N-PSDB; AAQ50452.
XX
XX New protein L binding light chains of all immunoglobulin classes
PT - for binding purifying and identifying immunoglobulin, also
PT related DNA, vectors and host cells
XX
PS Claim 1; Page 36; 71pp; English.
XX
XX The protein (Protein L) is capable of binding to immunoglobulin G
CC light chains. It is useful for binding, separating (purifying) and
CC identifying immunoglobulin and for removing immunoglobulin molecules
CC from serum. Hybrid proteins of the L protein can bind all human
CC immunoglobulin classes and many immunoglobulins from other species.
CC They are highly soluble and retain their binding activity at high
CC temperatures over a pH range of 3-10. They can be immobilised
CC without loss of activity.
XX
XX Sequence 305 AA;
XX
Query Match 100.0%; Score 1565; DB 14; Length 305;
Best Local Similarity 100.0%; Pred. No. 2.2e-110;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AVENKEETPETDSEEVTKANLIFANGSTQTAFFKGTFEKATSEAYADTLKKN 60
DB 1 AVENKEETPETDSEEVTKANLIFANGSTQTAFFKGTFEKATSEAYADTLKKN 60
QY 61 GEYTVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTAFFKGTFEATAE 120
DB 61 GEYTVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTAFFKGTFEATAE 120
QY 121 AYRYADALKKNGEYTVADVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTA 180
DB 121 AYRYADALKKNGEYTVADVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTA 180
QY 181 EFKGTFEATAEAYRYADLLAKENGYTVADVADKGYTLNKFAGKEKTPPEPKKEVTIKA 240
DB 181 EFKGTFEATAEAYRYADLLAKENGYTVADVADKGYTLNKFAGKEKTPPEPKKEVTIKA 240
QY 241 NLIYADGKTQTAFFKGTFAEATAEAYRYADLLAKENGYTVADLEDGGYTNIRFAGKKVD 300
DB 241 NLIYADGKTQTAFFKGTFAEATAEAYRYADLLAKENGYTVADLEDGGYTNIRFAGKKVD 300
QY 301 EKPEE 305
DB 301 EKPEE 305

```

RESULT 2

AA42994
ID AA42994 standard; Protein; 434 AA.

XX
AC AA42994;

XX 16-MAY-1994 (first entry)

XX Sequence encoding immunoglobulin light chain binding protein.

XX Immunoglobulin; light chain; binding; identification; purification;

XX separation; ss.

XX E. coli L392/pHDLG, DSM 7055.

OS

```

XX Location/Qualifiers
FH 5..305
FT /label= B1 immunoglobulin light chain binding
FT domain.
FT 81..305
FT /label= B2 immunoglobulin light chain binding
FT domain.
FT 153..305
FT /label= B3 immunoglobulin light chain binding
FT domain.
FT 225..305
FT /label= B4 immunoglobulin light chain binding
FT domain.
FT 297..305
FT /label= B5 immunoglobulin light chain binding
FT domain.
FT 309..434
FT /label= C1 immunoglobulin heavy chain binding
FT domain.
FT 364..434
FT /label= D intermediate immunoglobulin heavy
FT chain binding domain.
FT 379..434
FT /label= C2 immunoglobulin heavy chain binding
FT domain.
XX
XX WO9322342-A.
XX
XX 11-NOV-1993.
XX
XX 28-APR-1993; 93WO-SE00375.
XX
XX 28-APR-1992; 92SE-0001331.
XX
XX (HIGH-) HIGHTECH RECEPTOR AB.
XX
XX Bjoerck L, Sjoerbring U;
XX WPI; 1993-368722/46.
XX DR P-PSDB; AA42994.
XX
XX New protein L binding light chains of all immunoglobulin classes
PT - for binding purifying and identifying immunoglobulin, also
PT related DNA, vectors and host cells
XX
XX Claim 6; Page 39-40; 71pp; English.
XX
XX Protein L (AA42993) is capable of binding to immunoglobulin G light
CC chains. It is useful for binding, separating (purifying) and
CC identifying immunoglobulin and for removing immunoglobulin molecules
CC from serum. This is the coding sequence of one hybrid protein of
CC the L protein. The hybrid proteins can bind all human
CC immunoglobulin classes and many immunoglobulins from other species.
CC They are highly soluble and retain their binding activity at high
CC temperatures over a pH range of 3-10. They can be immobilised
CC without loss of activity.
XX
XX Sequence 434 AA;
XX
Query Match 100.0%; Score 1565; DB 14; Length 434;
Best Local Similarity 100.0%; Pred. No. 3.4e-110;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AVENKEETPETDSEEVTKANLIFANGSTQTAFFKGTFEKATSEAYADTLKKN 60
DB 1 AVENKEETPETDSEEVTKANLIFANGSTQTAFFKGTFEKATSEAYADTLKKN 60
QY 61 GEYTVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTAFFKGTFEATAE 120
DB 61 GEYTVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTAFFKGTFEATAE 120
QY 121 AYRYADALKKNGEYTVADVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTA 180

```


|||||
Db 121 AYRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 180
QY 181 EFKGTFEETAEAYRYADLLAKENGYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKA 240
Db 181 EFKGTFEETAEAYRYADLLAKENGYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKA 240
QY 241 NLIYADGKTQTAEFKGTFAETAEAYRYADLLAKENGYTADLEDGGYTNIRFAGKKVD 300
Db 241 NLIYADGKTQTAEFKGTFAETAEAYRYADLLAKENGYTADLEDGGYTNIRFAGKKVD 300
QY 301 EKPEE 305
Db 301 EKPEE 305
RESULT 3
AAB31372
ID AAB31372 standard; Protein; 467 AA.
XX
AC AAB31372;
XX
DT 20-APR-2001 (first entry)
XX
DE Amino acid sequence of protein L/CBD cex/ER retaining peptide fusion.
KW Protein production; food processing; protein antibiotic; feed enzyme;
KW protein L: CBD cex protein; cell signal peptide.
XX
OS Synthetic.
XX
PN W0200077174-A1.
XX
PD 21-DEC-2000.
XX
PF 07-JUN-2000; 2000WO-IL00330.
XX
PR 10-JUN-1999; 99US-0329234.
XX
PA (CBDT-) CBD TECHNOLOGIES LTD.
PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
PI Shani Z, Shoseyov O;
PI
DR WPI: 2001-112219/12.
DR N-PSDB; AAF24730.
XX
XX
PT Expressing and isolating recombinant protein in a plant, useful for
PT producing large quantities of recombinant proteins, by expressing a
PT fusion protein including a cellulose binding peptide fused to a
PT recombinant protein -
XX
PS Example: Fig 2a; 87pp; English.
XX
CC The specification describes a method for expressing and isolating
CC a recombinant protein in a plant. The method comprising expressing a
CC fusion protein including the recombinant protein and a cellulose
CC binding peptide fused to it, where the fusion protein is
CC compartmentalised and sequestered within plant cells, plant derived
CC tissue or cultured plant cells. The method is useful for obtaining large
CC quantities of the recombinant proteins and protein products in a simple
CC and cost-effective manner. Recombinant proteins may be used commercially,
CC such as in the food processing industry, e.g. glucoamylases and glucose
CC isomerases are used for converting starch to high fructose corn syrup,
CC proteinases for the hydrolysis of high molecular weight proteins and in
CC manufacturing leather or alcoholic beverages, pectinesterases for
CC pectin hydrolysis in food industry, lipases for cleaving ester linkage
CC in triglycerides, and for effluent treatment. The recombinant proteins
CC may further be used to produce protein antibiotics, which can be used
CC in healing processes, and to produce animal feed enzymes. The present
CC sequence represents a fusion protein of the invention, and comprises a
CC fusion of a cell signal peptide, protein L, CBD cex, and an endoplasmic
CC reticulum retaining peptide.

XX
SQ Sequence 467 AA;
Query Match 98.7%; Score 1544.5; DB 22; Length 467;
Best Local Similarity 99.3%; Pred. No. 1.3e-108;
Matches 303; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 AVENKEETPTPTDSEEEVTIKANLIFANGSTQTAEFKGTFEKATSEAYAYADTLKKDN 60
|||||
Db 40 AVENKEETPTPTDSEEEVTIKANLIFANGSTQTAEFKGTFEKATSEAYAYADTLKKDN 99
|||||
QY 61 GEYTVADVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTAEFKGTFEETAE 120
|||||
Db 100 GEYTVADVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTAEFKGTFEETAE 159
|||||
QY 121 AYRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 180
|||||
Db 160 AYRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 219
|||||
QY 181 EFKGTFEETAEAYRYADLL-ARENGKYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIK 239
|||||
Db 220 EFKGTFEETAEAYRYADLLAAKENGKYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIK 279
|||||
QY 240 ANLIYADGKTQTAEFKGTFAETAEAYRYADLLAKENGYTADLEDGGYTNIRFAGKKV 299
|||||
Db 280 ANLIYADGKTQTAEFKGTFAETAEAYRYADLLAKENGYTADLEDGGYTNIRFAGKKV 339
|||||
QY 300 DEKPE 304
Db 340 DEKPE 344
RESULT 4
AAB10432
ID AAB10432 standard; Protein; 367 AA.
XX
AC AAB10432;
XX
DT 01-DEC-2000 (first entry)
XX
DE Expression vector pSEX1114 protein G.
XX
KW Expression vector; antibody binding protein; monoclonal antibody; Neo-R;
KW B lymphocyte; myeloma cell; hybridoma; protein G; bla protein.
XX
OS Synthetic.
XX
PN DE19900635-A1.
XX
PD 13-JUL-2000.
XX
PF 11-JAN-1999; 99DE-1000635.
XX
PR 11-JAN-1999; 99DE-1000635.
XX
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI Breitling F, Poustka A, Moldenhauer G;
XX
DR WPI: 2000-499832/45.
DR N-PSDB; AAA71428.
XX
PT Selecting monoclonal antibodies, by expressing them on the surface of
PT hybridomas attached to antibody-binding protein, then reaction with
PT antibody library -
XX
PS Claim 16; Fig 1; 22pp; German.
XX
CC This invention describes a novel method for the selection of monoclonal
CC antibodies (MAB) which comprises (i) fusing B lymphocytes with myeloma
CC cells to produce antibody-producing hybridomas such that the antibodies
CC are presented at the surface of the hybridomas by an antibody-binding
CC protein (I); and (ii) binding the antibody to antigens (Ag). The

PD 02-OCT-1997.
 XX PF - 26-MAR-1997; 97WO-AU00194.
 XX
 XX 27-FEB-1997; 97AU-0005375.
 PR 26-MAR-1996; 96AU-0008951.
 XX
 PA (AMRA-) AMRAD OPERATIONS PTY LTD.
 XX
 XX Koentgen F, Suess GM, Tarlington DM, Treutlein HR;
 PI WPI; 1997-489572/45.
 XX N-PSDB; AAT91589.
 DR
 XX New catalytic antibody precursors - comprising a B-cell surface
 PT molecule binding portion which can induce B-cell mitogenesis
 PT
 XX Example 11; Page 66-68; 109pp; English.
 PS
 XX This protein comprises growth factor TLHL, where L is the
 CC immunoglobulin binding entity from Peptostreptococcus magnus, H is
 CC residues 42-62 of hen egg lysozyme, and T represents the variable
 CC kappa light chain (see AW32483) from human myeloma protein LEN. It
 CC was expressed in E. coli DH10B cells utilising a DNA construct (see
 CC AAT91589) produced from LHL (see AAT91986), kappa (see AAT91590) and
 CC synthetic linker oligonucleotides, and was purified from total
 CC bacterial lysate via a strep-tag. The linker contains a cleavage
 CC site for tobacco etch virus (TEV) protease. TLHL was designed so
 CC that the kappa portion of the protein is cleaved by TEV protease so
 CC into 'T' and 'LHL' moieties. Production of catalytic antibodies to
 CC a specific antigen comprises administering to an animal a growth
 CC factor comprising an antigen capable of interacting with a B cell
 CC bound catalytic antibody. The antigen is fused to a B cell surface
 CC molecule binding protein for the antigen to be cleaved and for the
 CC remainder of the molecule to induce B cell mitogenesis (claimed).
 CC LHL crosslinks with surface immunoglobulin on B cells. This induces
 CC B cell activation and blast formation. Internalisation and
 CC processing of LHL leads to the presentation of the H peptide on
 CC MHC II. T cell recognition of MHC II with the H peptide signals the
 CC activated B cell to proliferate and undergo antibody class switching
 CC and secretion. The catalytic antibodies generated by the process
 CC can have diagnostic and therapeutic applications.
 XX
 SQ Sequence 342 AA;
 Query Match 39.5%; Score 618; DB 18; Length 342;
 Best Local Similarity 75.9%; Pred. No. 6e-39;
 Matches 126; Conservative 5; Mismatches 15; Indels 20; Gaps 1;
 QY 16 SEEVITKANLIFANGSTQTAEPFGTPEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 75
 Db 174 SAEVITKANLIFANGSTQTAEPFGTPEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 233
 QY 76 IKFAGKEKTPPEE-----PKEEVITKANLIYADGKTQTAEFGKTFE 115
 Db 234 IKFAGKEATNRNTDGDSTDYGILQINSRWGLTSAEEVITKANLIFANGSTQTAEPFGKTFE 293
 QY 116 EATAEAYRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPK 161
 Db 294 KATSEAYAYADTLKKDNGEYTVDVADKGYTLNKFAGKESAWRHPQ 339
 RESULT 9
 ID AAY06909
 XX AAY06909 standard; Protein; 342 AA.
 AC AAY06909;
 XX
 XX 01-JUL-1999 (first entry)
 XX TLHL amino acid sequence.
 XX Growth factor precursor; B-cell surface; T cell surface; Cab; hepatitis;

catalytic antibody; immunoglobulin; tumour necrosis factor; influenza;
 rheumatoid arthritis; toxic shock syndrome; viral docking receptor;
 human immune virus; tumour-specific antigen; amyloid plaque; myeloma;
 Alzheimer's disease; IGE; allergy; asthma; drug detoxification;
 autoimmune; inflammatory disease; gene therapy; protein L; P. magnus;
 hen egg lysozyme; HEL; LHL; TLHL.
 XX
 XX Peptostreptococcus magnus.
 OS
 OS Gallus sp.
 XX
 PN WO9915563-A1.
 XX
 XX 01-APR-1999.
 PD
 XX
 XX 18-SEP-1998; 98WO-AU00783.
 PF
 XX 19-SEP-1997; 97AU-0009306.
 PR
 XX (AMRA-) AMRAD OPERATIONS PTY LTD.
 PA
 XX Koentgen F, Suess GM, Tarlington DM, Treutlein HR;
 PI WPI; 1999-244394/20.
 XX N-PSDB; AAX34592.
 DR
 XX Growth factor precursor cleaved by antigen-specific catalytic
 PT antibody
 PT
 XX Example 11; Page 67-69; 101pp; English.
 PS
 XX The invention relates to a growth factor precursor that comprises B-cell
 CC surface binding part, T cell surface binding part, antigen cleavable
 CC by a catalytic antibody (CAB); and a peptide comprising heavy and light
 CC chains of immunoglobulin. When the antigen is cleaved the B cell surface
 CC part can interact with its target. The growth factor precursors are used
 CC to select B cells that produce Ag-specific CAB, and to generate CAB from
 CC such cells (by inducing mitogenesis, caused by the growth factor released
 CC by specific cleavage). The Ag-specific CAB can be directed against, e.g.
 CC tumour necrosis factor (treatment of rheumatoid arthritis or toxic shock
 CC syndrome); viral docking receptors (treatment of human immune virus,
 CC hepatitis and influenza infections); tumour-specific antigens; amyloid
 CC plaque (treatment of Alzheimer's disease or myeloma) or IGE (treatment of
 CC allergies such as asthma). CAB may also be used for drug detoxification,
 CC to treat autoimmune or inflammatory diseases and to eliminate
 CC environmental or industrial pollutants, such as plastics and petroleum.
 CC Particularly the growth factor precursors are produced by delivering the
 CC corresponding nucleic acid in a viral or other gene therapy vector. The
 CC present sequence represents the amino acid sequence of TLHL. The LHL is a
 CC growth factor comprising H flanked by two L molecules where L is a B cell
 CC surface molecule binding portion from protein L of P. magnus; H is a T
 CC cell surface molecule binding portion from hen egg lysozyme (HEL). TLHL
 CC is LHL linked to kappa chain via TEV (tobacco etch virus) sensitive
 CC peptide and g attached to N terminus region.
 XX
 SQ Sequence 342 AA;
 Query Match 39.5%; Score 618; DB 20; Length 342;
 Best Local Similarity 75.9%; Pred. No. 6e-39;
 Matches 126; Conservative 5; Mismatches 15; Indels 20; Gaps 1;
 QY 16 SEEVITKANLIFANGSTQTAEPFGTPEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 75
 Db 174 SAEVITKANLIFANGSTQTAEPFGTPEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 233
 QY 76 IKFAGKEKTPPEE-----PKEEVITKANLIYADGKTQTAEFGKTFE 115
 Db 234 IKFAGKEATNRNTDGDSTDYGILQINSRWGLTSAEEVITKANLIFANGSTQTAEPFGKTFE 293
 QY 116 EATAEAYRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPK 161
 Db 294 KATSEAYAYADTLKKDNGEYTVDVADKGYTLNKFAGKESAWRHPQ 339

RESULT 10

AAW32486
 ID AAW32486 standard; Protein; 178 AA.
 XX
 AC AAW32486;
 XX
 DT 27-MAR-1998 (first entry)
 XX
 DE Growth factor LHL-omp (catalytic antibody precursor).
 XX
 DE Catalytic antibody; growth factor; B-cell mitogenesis; LHL-omp;
 KW L protein; hen egg lysozyme; HEL.
 XX
 OS Chimeric - Peptostreptococcus magnus.
 OS Chimeric - Gallus sp.
 OS Chimeric - Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Peptide 2..9
 FT /note= "FLAG epitope"
 FT 170..178
 FT /note= "strept-tag"
 XX
 PN W09735887-A1.
 XX
 PD 02-OCT-1997.
 XX
 XX 26-MAR-1997; 97WO-AU00194.
 XX
 PR 27-FEB-1997; 97AU-0005375.
 PR 26-MAR-1996; 96AU-0008951.
 XX
 PA (AMRA-) AMRAD OPERATIONS PTY LTD.
 XX
 PI Koentgen F, Suess GM, Tarlinton DM, Treutlein HR;
 XX
 DR WPI; 1997-489572/45.
 DR N-PSDB; AAT91591.
 XX
 DR New catalytic antibody precursors - comprising a B-cell surface
 PT molecule binding portion which can induce B-cell mitogenesis
 XX
 PS Example 27; Page 76; 109pp; English.
 XX
 CC This polypeptide comprises growth factor LHL-omp, where L is the
 CC immunoglobulin binding entity from Peptostreptococcus magnus and H
 CC is residues 42-62 of hen egg lysozyme. It was generated from
 CC LHL.seq (see AAW32481) such that the E. coli ompA signal sequence
 CC (generated by purifying LHL.seq from whole bacterial host cells) was
 CC eliminated. The N-terminal omp peptide found in LHL (see AAW32479),
 CC LHL.seq (see AAW32481) and TLHL (see AAW32482) induces multimerisation
 CC as demonstrated by a potentiation of biological activity as
 CC compared to recombinant LHL-omp. Production of catalytic antibodies
 CC to a specific antigen comprises administering to an animal a growth
 CC factor comprising an antigen capable of interacting with a B cell
 CC bound catalytic antibody. The antigen is fused to a B cell surface
 CC molecule binding protein for the antigen to be cleaved and for the
 CC remainder of the molecule to induce B cell mitogenesis (claimed).
 CC The observation that omp induces multimerisation allows the design
 CC of simpler molecules with the same desired biological function as
 CC LHL, TLHL and CATAB (see AAW32480). The catalytic antibodies
 CC generated by the process can have diagnostic and therapeutic
 CC applications.
 XX
 SQ Sequence 178 AA;
 Query Match 39.3%; Score 615; DB 18; Length 178;
 Best Local Similarity 76.2%; Pred. No. 4.5e-39;
 Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1;
 18 EVVTIKANLIFANGSTQAEFGKTFKATSEAYAYADTLKKDNGEYTVDVADKGYTLNIK 77
 Db 12 EVVTIKANLIFANGSTQAEFGKTFKATSEAYAYADTLKKDNGEYTVDVADKGYTLNIK 71

QY 78 FAGKEKTPPE-----PKEEVTIKANLIYADGKTQTAERKGTPEEA 117
 DB 72 FAGKEATNRNTDGSYGIQINSRWGGLTSAEVTIKANLIFANGSTQTAERKGTPEKA 131
 QY 118 TAEAYRYADALKKDNGEYTVDVADKGYTLNIKFAGKEKTPPEPK 161
 DB 132 TSEAYAYADTLKKDNGEYTVDVADKGYTLNIKFAGKESAWRHPQ 175
 RESULT 11
 AAY06913
 ID AAY06913 standard; Protein; 178 AA.
 XX
 AC AAY06913;
 XX
 DT 01-JUL-1999 (first entry)
 XX
 DE Amino acid sequence of a form of LHL growth factor (LHL.omp).
 XX
 KW Growth factor precursor; B-cell surface; T cell surface; CAB; hepatitis;
 KW catalytic antibody; immunoglobulin; tumour necrosis factor; influenza;
 KW rheumatoid arthritis; toxic shock syndrome; viral docking receptor;
 KW human immune virus; tumour-specific antigen; amyloid plaque; myeloma;
 KW Alzheimer's disease; IGE; allergy; asthma; drug detoxification;
 KW autoimmune; inflammatory disease; gene therapy; protein L; P. magnus;
 KW hen egg lysozyme; HEL; LHL.
 XX
 OS Peptostreptococcus magnus.
 OS Gallus sp.
 XX
 PN W09915563-A1.
 XX
 PD 01-APR-1999.
 XX
 XX 18-SEP-1998; 98WO-AU00783.
 XX
 PR 19-SEP-1997; 97AU-0009306.
 XX
 PA (AMRA-) AMRAD OPERATIONS PTY LTD.
 XX
 PI Koentgen F, Suess GM, Tarlinton DM, Treutlein HR;
 XX
 DR WPI; 1999-244394/20.
 DR N-PSDB; AAX34595.
 XX
 PT Growth factor precursor cleaved by antigen-specific catalytic
 PS antibody
 XX
 PS Examples; Page 74; 101pp; English.
 CC The invention relates to a growth factor precursor that comprises B-cell
 CC surface binding part, T cell surface binding part, antigen cleavable
 CC by a catalytic antibody (CAB); and a peptide comprising heavy and light
 CC chains of immunoglobulin. When the antigen is cleaved the B cell surface
 CC part can interact with its target. The growth factor precursors are used
 CC to select B cells that produce Ag-specific CAB, and to generate CAB from
 CC such cells (by inducing mitogenesis, caused by the growth factor released
 CC by specific cleavage). The Ag-specific CAB can be directed against, e.g.
 CC tumour necrosis factor (treatment of rheumatoid arthritis or toxic shock
 CC syndrome); viral docking receptors (treatment of human immune virus,
 CC hepatitis and influenza infections); tumour-specific antigens; amyloid
 CC plaque (treatment of Alzheimer's disease or myeloma) or IGE (treatment of
 CC allergies such as asthma). CAB may also be used for drug detoxification,
 CC to treat autoimmune or inflammatory diseases and to eliminate
 CC environmental or industrial pollutants, such as plastics and petroleum.
 CC Particularly the growth factor precursors are produced by delivering the
 CC corresponding nucleic acid in a viral or other gene therapy vector. The
 CC present sequence represents the amino acid sequence of the LHL growth
 CC factor carrying an ompA signal peptide (LHL.omp). L is a B cell surface
 CC molecule binding portion from protein L of P. magnus; H is a T cell
 CC surface molecule binding portion from hen egg lysozyme (HEL). LHL is a
 CC growth factor comprising H flanked by two L molecules.

XX SQ Sequence 178 AA;
 Query Match 39.3%; Score 615; DB 20; Length 178;
 Best Local Similarity 76.2%; Pred. No. 4.5e-39;
 Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1;
 QY 18 EEVTKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLNLIK 77
 Db 12 EEVTKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLNLIK 71
 QY 78 FAGKEKTPPEE-----PKKEVTIKANLIYADGKTQTAEFKGTFFEA 117
 Db 72 FAGKEATNRNTDGTGYILOINSRWGLTSABEVTIKANLIFANGSTQTAEFKGTFFKA 131
 QY 118 TAEAYRYADALKKDNGEYTVDVADKGYTLNLIK FAGKEKTPPEEPK 161
 Db 132 TSEAYAYADTLKKDNGEYTVDVADKGYTLNLIK FAGKESAWRHQ 175
 RESULT 12
 AAW32481
 ID AAW32481 standard; Protein; 198 AA.
 AC AAW32481;
 XX 27-MAR-1998 (first entry)
 DE Growth factor LHL.seq (catalytic antibody precursor).
 XX Catalytic antibody; growth factor; B-cell mitogenesis; LHL.seq;
 KW L protein; hen egg lysozyme; HEL.
 OS Chimeric - Peptostreptococcus magnus.
 OS Chimeric - Gallus sp.
 OS Chimeric - Synthetic.
 XX Key Location/Qualifiers
 FT Peptide 22..29
 FT /note= "FLAG epitope"
 FT Peptide 190..198
 FT /note= "strept-tag"
 XX WO9735887-A1.
 XX 02-OCT-1997.
 XX 26-MAR-1997; 97WO-AU00194.
 XX 27-FEB-1997; 97AU-0005375.
 XX 26-MAR-1996; 96AU-0008951.
 XX (AMRA-) AMRAD OPERATIONS PTY LTD.
 XX Koentgen F, Suess GM, Tarlington DM, Treutlein HR;
 WPI: 1997-489572/45.
 N-PSDB; AAT91588.
 XX New catalytic antibody precursors - comprising a B-cell surface
 molecule binding portion which can induce B-cell mitogenesis
 XX Example 3; Page 70-71; 109pp; English.
 CC This polypeptide comprises growth factor LHL.seq, where L is the
 CC immunoglobulin binding entity from Peptostreptococcus magnus and H
 CC is residues 42-62 of hen egg lysozyme. It is a modified form of
 CC LHL (see AAW32479) having an N-terminal FLAG epitope (see AAW32484) and
 CC a C-terminal strep-tag (see AAW32485) that facilitates purification.
 CC Production of catalytic antibodies to a specific antigen comprises
 CC administering to an animal a growth factor comprising an antigen
 CC capable of interacting with a B cell bound catalytic antibody. The
 CC antigen is fused to a B cell surface molecule binding protein for

the antigen to be cleaved and for the remainder of the molecule to
 induce B cell mitogenesis (claimed). LHL.seq has the same
 biological activity as LHL, which crosslinks with surface
 immunoglobulin on B cells. This induces B cell activation and blast
 formation. Internalisation and processing of LHL leads to the
 presentation of the H peptide on MHC II. T cell recognition of MHC
 II with the H peptide signals the activated B cell to proliferate
 CC and undergo antibody class switching and secretion. The catalytic
 CC antibodies generated by the process can have diagnostic and
 CC therapeutic applications.
 XX SQ Sequence 198 AA;
 Query Match 39.3%; Score 615; DB 18; Length 198;
 Best Local Similarity 76.2%; Pred. No. 5.1e-39;
 Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1;
 QY 18 EEVTKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLNLIK 77
 Db 32 EEVTKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLNLIK 91
 QY 78 FAGKEKTPPEE-----PKKEVTIKANLIYADGKTQTAEFKGTFFEA 117
 Db 92 FAGKEATNRNTDGTGYILOINSRWGLTSABEVTIKANLIFANGSTQTAEFKGTFFKA 151
 QY 118 TAEAYRYADALKKDNGEYTVDVADKGYTLNLIK FAGKEKTPPEEPK 161
 Db 152 TSEAYAYADTLKKDNGEYTVDVADKGYTLNLIK FAGKESAWRHQ 195
 RESULT 13
 AAY06910
 ID AAY06910 standard; Protein; 198 AA.
 AC AAY06910;
 XX 01-JUL-1999 (first entry)
 DE Amino acid sequence of a form of LHL growth factor (LHL.seq).
 XX Growth factor precursor; B-cell surface; T cell surface; CAB; hepatitis;
 KW catalytic antibody; immunoglobulin; tumour necrosis factor; influenza;
 KW rheumatoid arthritis; toxic shock syndrome; viral docking receptor;
 KW human immune virus; tumour-specific antigen; amyloid plaque; myeloma;
 KW Alzheimer's disease; IGE; allergy; asthma; drug detoxification;
 KW autoimmune; inflammatory disease; gene therapy; protein L; P. magnus;
 KW hen egg lysozyme; HEL; LHL; LHL.seq.
 XX Peptostreptococcus magnus.
 OS Gallus sp.
 XX WO9915563-A1.
 XX 01-APR-1999.
 XX 18-SEP-1998; 98WO-AU00783.
 XX 19-SEP-1997; 97AU-0009306.
 XX (AMRA-) AMRAD OPERATIONS PTY LTD.
 XX Koentgen F, Suess GM, Tarlington DM, Treutlein HR;
 WPI: 1999-244394/20.
 N-PSDB; AAX34593.
 XX Growth factor precursor cleaved by antigen-specific catalytic
 PT antibody
 XX Example 3; Page 70-71; 101pp; English.
 PS The invention relates to a growth factor precursor that comprises B-cell
 CC surface binding part, T cell surface binding part, antigen cleavable

CC by a catalytic antibody (CAB); and a peptide comprising heavy and light
 CC chains of immunoglobulin. When the antigen is cleaved the B cell surface
 CC part can interact with its target. The growth factor precursors are used
 CC to select B cells that produce Ag-specific Cab, and to generate CAB from
 CC such cells (by inducing mitogenesis, caused by the growth factor released
 CC by specific cleavage). The Ag-specific CAB can be directed against, e.g.
 CC tumour necrosis factor (treatment of rheumatoid arthritis or toxic shock
 CC syndrome); viral docking receptors (treatment of human immune virus,
 CC hepatitis and influenza infections); tumour-specific antigens; amyloid
 CC plaque (treatment of Alzheimer's disease or myeloma) or IGF (treatment of
 CC allergies such as asthma). CAB may also be used for drug detoxification,
 CC to treat autoimmune or inflammatory diseases and to eliminate
 CC environmental or industrial pollutants, such as plastics and petroleum.
 CC Particularly the growth factor precursors are produced by delivering the
 CC corresponding nucleic acid in a viral or other gene therapy vector. The
 CC present sequence represents the amino acid sequence of the LHL growth
 CC factor carrying a N-terminal FLAG epitope and a C-terminal strep tag.
 CC L is a B cell surface molecule binding portion from protein L of
 CC P. magnus; H is a T cell surface molecule binding portion from hen egg
 CC lysozyme (HEL). LHL is a growth factor comprising H flanked by two L
 CC molecules.

XX Sequence 198 AA;

Query Match 39.3%; Score 615; DB 20; Length 198;
 Best Local Similarity 76.2%; Pred. No. 5.1e-39;
 Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1;

QY 18 EFTVTKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKDNGEYTVVDVADKGYTLNIK 77
 DB 32 EFTVTKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKDNGEYTVVDVADKGYTLNIK 91
 QY 78 FAGKEKTPPE-----PKEEVTKANLIYADGKTOTAEFGKTFEEA 117
 DB 92 FAGKEATNRNTDGTGYILQINSRWGLTSAEEVTKANLIFANGSTQTAEFKGTPEKA 151
 QY 118 TABAYRAYADALKKDNGEYTVVDVADKGYTLNIKFAKGEKTPPEPK 161
 DB 152 TSEAYAYADTLKKDNGEYTVVDVADKGYTLNIKFAKGESAWRHQ 195

RESULT 14

ID AAW32480
 XX AAW32480 standard; Protein; 495 AA.

AC AAW32480;

XX 27-MAR-1998 (first entry)

DE Growth factor CATAB-TEV (catalytic antibody precursor).

XX Catalytic antibody; growth factor; B-cell mitogenesis; LHL;
 KW L protein; hen egg lysozyme; kappa light chain.

XX Chimeric - Peptostreptococcus magnus.

OS Chimeric - Gallus sp.

OS Chimeric - Homo sapiens.

PN W09735887-AL.

XX 02-OCT-1997.

XX 26-MAR-1997; 97WO-AU00194.

XX 27-FEB-1997; 97AU-0005375.

PR 26-MAR-1996; 96AU-0008951.

XX (AMRA-) AMRAD OPERATIONS PTY LTD.

XX Koentgen F, Suess GM, Tarlinton DM, Treutlein HR;

XX WPI; 1997-489572/45.

DR N-PSDB; AAT91587.

XX

PT New catalytic antibody precursors - comprising a B-cell surface
 molecule binding portion which can induce B-cell mitogenesis

PS Claim 12; Page 61-63; 109pp; English.

XX

CC This polypeptide comprises CATAB-TEV, a synthetic growth factor
 CC composed of: (a) TLHL (see AAW91589), where L is the immunoglobulin
 CC binding entity of Peptostreptococcus magnus, H is residues 42-62 of
 CC hen egg lysozyme, and T represents the variable kappa light chain
 CC (see AAW32483) from human myeloma protein LEN; and (b) an additional,
 CC C-terminal kappa protein, the elements of CATAB-TEV being joined by
 CC linkers containing recognition sites for tobacco etch virus (TEV)
 CC protease. CATAB-TEV was designed so that the kappa portions of the
 CC protein are removed by TEV protease to release 'T' and 'LHL'
 CC moieties. Production of catalytic antibodies to a specific antigen
 CC comprises administering to an animal a growth factor comprising an
 CC antigen capable of interacting with a B cell bound catalytic
 CC antibody. The antigen is fused to a B cell surface molecule
 CC binding protein for the antigen to be cleaved and for the remainder
 CC of the molecule to induce B cell mitogenesis (claimed). LHL (see
 CC AAW32479) crosslinks with surface immunoglobulin on B cells. This
 CC induces B cell activation and blast formation. Internalisation
 CC and processing of LHL leads to the presentation of the H peptide on
 CC MHC II. T cell recognition of MHC II with the H peptide signals the
 CC activated B cell to proliferate and undergo antibody class switching
 CC and secretion. The catalytic antibodies generated by the process
 CC can have diagnostic and therapeutic applications.

XX Sequence 495 AA;

Query Match 39.2%; Score 614; DB 18; Length 495;

Best Local Similarity 78.6%; Pred. No. 1.9e-38;

Matches 125; Conservative 4; Mismatches 10; Indels 20; Gaps 1;

QY 16 SEBEVTIKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKDNGEYTVVDVADKGYTLN 75

DB 174 SAEVETIKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKDNGEYTVVDVADKGYTLN 233

QY 76 IKFAGKEKTPPE-----PKEEVTKANLIYADGKTOTAEFGKTFE 115
 DB 234 IKFAGKEATNRNTDGTGYILQINSRWGLTSAEEVTKANLIFANGSTQTAEFKGTPE 293

QY 116 EATAEAYRAYADALKKDNGEYTVVDVADKGYTLNIKFAKGE 154

DB 294 KATSEAYAYADTLKKDNGEYTVVDVADKGYTLNIKFAKGE 332

RESULT 15

ID AAY06908

XX AAY06908 standard; Protein; 495 AA.

AC AAY06908;

XX 01-JUL-1999 (first entry)

DE CATAB-TEV aminoacid sequence.

XX Growth factor precursor; B-cell surface; T cell surface; CAB; hepatitis;
 KW catalytic antibody; immunoglobulin; tumour necrosis factor; influenza;
 KW rheumatoid arthritis; toxic shock syndrome; viral docking receptor;
 KW human immune virus; tumour-specific antigen; amyloid plaque; myeloma;
 KW Alzheimer's disease; IGE; allergy; asthma; drug detoxification;
 KW autoimmune; inflammatory disease; gene therapy; protein L; P. magnus;
 KW hen egg lysozyme; HEL; LHL; CATAB-TEV; tobacco etch virus; TLHL.

OS Peptostreptococcus magnus.

OS Gallus sp.

XX W09915563-AL.

XX 01-APR-1999.

PD

XX

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:25:26 ; Search time 10.571 Seconds
(without alignments)
1589.653 Million cell updates/sec

Title: US-08-325-278B-3

Perfect score: 2235

Sequence: 1 AVENKETPTETDSEEV.....GVDGVWYDDATKTFVTM 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	708	31.7	593	1 SPG2_STRSP	P19009 streptococ
2	683	30.6	448	1 SPG1_STRSP	P06654 streptococ
3	188.5	8.4	1183	1 CNA_STAAU	Q53654 staphylococ
4	156	7.0	915	1 NFM_HUMAN	P07197 homo sapien
5	155	6.9	555	1 EDCB_DAUCA	P20075 daucus caro
6	151.5	6.8	1637	1 MRSP_STAAU	P80544 staphylococ
7	149	6.7	1433	1 SUBE_BACSU	P16397 bacillus su
8	148.5	6.6	2647	1 ABP2_HUMAN	P21333 homo sapien
9	148	6.6	1233	1 YFI6_YEAST	P43597 saccharomyc
10	146.5	6.6	1337	1 DEXT_STRDO	P39653 streptococ
11	144	6.4	2334	1 WAPA_BACSU	Q07833 bacillus su
12	143.5	6.4	848	1 NFM_MOUSE	P08553 mus musculu
13	142.5	6.4	814	1 SLAI_BACAN	P49051 bacillus an
14	141.5	6.3	845	1 NFM_RAT	P12839 rattus norv
15	141	6.3	2459	1 MAPB_MOUSE	P15205 rattus norv
16	138.5	6.2	2464	1 MAPB_MOUSE	P14873 mus musculu
17	138	6.2	2468	1 MAPB_HUMAN	P46821 homo sapien
18	136.5	6.1	1500	1 SP55_STRGN	P16952 streptococ
19	134.5	6.0	1940	1 MYH3_CHICK	P02565 gallus gall
20	134	6.0	1935	1 MYSS_CYPCA	Q90339 cyprinus ca
21	133.5	6.0	810	1 NFM_BOVIN	Q07788 bos taurus
22	133	6.0	505	1 FLJB_SALTY	P52616 salmonella
23	132.5	5.9	599	1 TBB2_NEIMB	Q06988 neseria m
24	131.5	5.9	857	1 NFM_CHICK	P16053 gallus gall
25	131.5	5.9	1018	1 FNBA_STAAU	P14738 staphylococ
26	131.5	5.9	2116	1 MYB2_DICDI	P08799 dictyosteli
27	131	5.9	6669	1 NEBU_HUMAN	P20929 homo sapien
28	130	5.8	1277	1 IF3X_YEAST	Q03690 saccharomyc
29	129.5	5.8	1104	1 COLA_CLOPE	P43153 clostridium
30	129.5	5.8	1028	1 MYHA_CLOPE	Q28441 cryptotagus
31	129	5.8	1176	1 SLAP_BACSH	P38537 bacillus sp
32	127	5.7	929	1 CAIC_NOTVI	Q91145 notophthalm
33	125.5	5.6	827	1 CSG_HALVO	P25062 halobacteri

ALIGNMENTS

```
RESULT 1
SPG2_STRSP
ID SPG2_STRSP STANDARD; PRT; 593 AA.
AC PI9909;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Immunoglobulin G binding protein G precursor (IGG binding protein G).
GN SPG.
OS Streptococcus sp. (strain G148).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1306;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G148;
RX MEDLINE=88029445; PubMed=3665928;
RA Olsson A., Eliasson M., Guss B., Nilsson B., Hellman U.,
RA Lindberg M., Uhlen M.;
RT "Structure and evolution of the repetitive gene encoding
RT streptococcal protein G.";
RL Eur. J. Biochem. 168:319-324(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GX7805;
RX MEDLINE=88015586; PubMed=3658689;
RA Filpula D., Alexander P., Fahnestock S.R.;
RT "Nucleotide sequence of the protein G gene from Streptococcus GX7805,
RT and comparison to previously reported sequences.";
RL Nucleic Acids Res. 15:7210-7210(1987).
RN [3]
RP SEQUENCE OF 114-593 FROM N.A.
RC STRAIN=G148;
RX MEDLINE=86300657; PubMed=3017704;
RA Guss B., Eliasson M., Olsson A., Uhlen M., Frej A.-K., Joernvall H.,
RA Flock J.-I., Lindberg M.;
RT "Structure of the IgG-binding regions of streptococcal protein G.";
RL EMBO J. 5:1567-1575(1986).
RN [4]
RP STRUCTURE BY NMR OF 371-427.
RC STRAIN=G148;
RX MEDLINE=98290449; PubMed=9628485;
RA Malakauskas S.M., Mayo S.L.;
RT "Design, structure and stability of a hyperthermophilic protein
RT variant.";
RL Nat. Struct. Biol. 5:470-475(1998).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

P06546 bacillus br
P54938 oryctolagus
Q08372 plasmodium
Q02113 bacillus su
P09405 mus musculu
P08198 halobacteri
P38538 bacillus br
P16371 lactococcus
P15292 lactococcus
Q99715 homo sapien
P06180 xenopus lae
P43261 escherichia

34 125.5 5.6 1053 1 SLPM_BACBR
35 125 5.6 644 1 NFM_RABIT
36 124.5 5.6 3135 1 Q230_PLAFO
37 124 5.5 705 1 CWBA_BACSU
38 124 5.5 706 1 NUCL_MOUSE
39 124 5.5 852 1 CSG_HALHA
40 124 5.5 1116 1 SLPH_BACBR
41 124 5.5 1902 1 P1P_LACLC
42 124 5.5 1902 1 P3P_LACLC
43 124 5.5 3063 1 CAIC_HUMAN
44 123.5 5.5 589 1 HIBN_XENLA
45 123.5 5.5 934 1 EAE_ECO57


```
DR PDB; 1IGC; 03-JUN-95.
DR PDB; 1IGD; 01-NOV-94.
DR PDB; 2IGD; 29-JUL-98.
DR PDB; 1FCC; 20-JUL-95.
DR PDB; 1GB1; 15-APR-93.
DR InterPro; IPR002988; GA.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000724; Igg_bind_B.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF01378; Igg_binding_B; 2.
DR PRINTS; PR00015; GPOSANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW Igg-binding protein; Repeat; Transmembrane; Cell wall; Signal;
FT SIGNAL 1 33
FT CHAIN 34 448 IMMUNOGLOBULIN G BINDING PROTEIN G.
FT DOMAIN 34 422 EXTRACELLULAR.
FT TRANSMEM 423 443
FT DOMAIN 444 448 CYTOPLASMIC.
FT DOMAIN 69 111 ALA-RICH.
FT DOMAIN 104 215 2 X 37 AA REPEATS.
FT REPEAT 104 140 1-1.
FT REPEAT 179 215 1-2.
FT DOMAIN 228 352 2 X 55 AA REPEATS.
FT REPEAT 228 282 2-1.
FT REPEAT 298 352 2-2.
FT DOMAIN 386 410 5 X 5 AA REPEATS OF [DE]-D-A-K-K.
FT DOMAIN 414 419 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 448 AA; 47567 MW; A0759060C8F9E6CA CRC64;

Query Match 30.6%; Score 683; DB 1; Length 448;
Best Local Similarity 53.2%; Pred. No. 1.9e-30;
Matches 176; Conservative 15; Mismatches 92; Indels 48; Gaps 9;

QY 130 KMGKYTVVDKGYTLNIFAGKEKTPPEKPEVITKANLIYAD-GKTOTAFFKGTFFEE 188
Db 45 RGGELT-----NLNGSETTLALRNEESATDLTAADVADTVAAAAENAGA 92
QY 189 ATAERYADLLA-----KENGKYTVVDKGYTLNIFAGKEKTPPEKPEVITKAN 241
Db 93 AAWEAAADALAKAKADALKEFNKYGVSDYKYNLNNAKTVEGIDQLQAVVESAKKAR 152
QY 242 LIYA-DG-----KTOTAFFKGTAAEAYRYAD-----LLAKENGKYTADLEDGGYITNI 292
Db 153 ISEATDGLSDFLSKQTP-----AEDTVKSIELAEAKVLANRELDKYGV-----SDYHKNL 202
QY 293 RFAGKKYVDEKPE-----EPMDTYKLINGKTLKGETTTEAVDAATAEKVKQYAND 343
Db 203 INNAKTVEGVKELIDEILAAALPKTDYKYLINGKTLKGETTTEAVDAATAEKVKQYAND 262
QY 344 NGVDGEWYDDATKFTFTVEKPEVIDASELTPAVTYYKLIVNGKTLKGETTTTAVDAETA 403
Db 263 NGVDGEWYDDATKFTFTVEKPEVIDASELTPAVTYYKLIVNGKTLKGETTTTAVDAETA 322
QY 404 EKAFKQYANDNGVDGVWYDDATKFTFTVTEM 434
Db 323 EKAFKQYANDNGVDGVWYDDATKFTFTVTEM 353

RESULT 3
CNA_STAAU
ID CNA_STAAU STANDARD; PRT; 1183 AA.
AC Q53634;
DT 15-DEC-1998 (Rel. 37, Created)
DF 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen adhesin precursor.
GN CNA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.

NCBI_TaxID=1280;
[1]
SEQUENCE FROM N.A.
RX STRAIN=FDA 574;
RX MEDLINE=92165839; PubMed=1311320;
Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
Lindberg M., Hoeoek M.;
"molecular characterization and expression of a gene encoding a
Staphylococcus aureus collagen adhesin.";
J. Biol. Chem. 267:4766-4772(1992).
[2]
ERRATUM.
Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
Lindberg M., Hoeoek M.;
J. Biol. Chem. 269:11672-11672(1994).
[3]
COLLAGEN-BINDING DOMAIN.
RX STRAIN=FDA 574;
RX MEDLINE=94032261; PubMed=8218209;
Patti J.M., Boles J.O., Hoeoek M.;
"Identification and biochemical characterization of the ligand
binding domain of the collagen adhesin from Staphylococcus aureus.";
Biochemistry 32:11428-11435(1993).
[4]
RX X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
MEDLINE=97475225; PubMed=9334749;
Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
Moore D., Jin L., Schneider A., Delucas L.J., Hoeoek M.,
Narayana S.V.L.;
"Structure of the collagen-binding domain from a Staphylococcus
aureus adhesin.";
Nat. Struct. Biol. 4:833-838(1997).
-!- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO
COLLAGEN-CONTAINING SUBSTRATA.
-!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
-!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
IN THE REGION OF THE MEMBRANE ANCHOR.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
-----
EMBL; M61736; AAA20874.1; -.
FDB; IAMX; 24-JUN-98.
InterPro; IPR001899; Gram_pos_anchor.
PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
Signal; Repeat; Transmembrane; Cell wall; 3D-structure.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 1183 COLLAGEN ADHESIN.
FT DOMAIN 30 1157 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1158 1177 MEMBRANE ANCHOR (POTENTIAL).
FT DOMAIN 1178 1183 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 151 318 COLLAGEN-BINDING.
FT DOMAIN 533 1093 3 X 187 AA APPROXIMATE TANDEM REPEATS.
FT DOMAIN 1093 1157 LYS/PRO-RICH (CELL WALL-SPANNING).
FT DOMAIN 1151 1156 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
FT REPEAT 533 719 B1.
FT REPEAT 720 906 B2.
FT REPEAT 907 1093 B3.
FT SEQUENCE 1183 AA; 133066 MW; B6A1CC072E575D76 CRC64;

Query Match 8.4%; Score 188.5; DB 1; Length 1183;
Best Local Similarity 23.8%; Pred. No. 0.0032;
Matches 136; Conservative 61; Mismatches 166; Indels 209; Gaps 38;

QY 2 VENKEETPETPETDSE-----EVTYKANLIFANG-STQTAFFKGTTEKAT 46
Db 525 VTNTEKPIETTSISGEKVDKQDQKRPKYSVN---LLANGKVKYTLTLD-----VT 574
```


QY 358 --TFTVTEKPEVIDASELTPAVTTTKLVINGKTLKGTGTTTKAVDAETAETAEKAFKQYANDNG 415
 Db 818 KGSREBEKGVINGLDLSPA-----DEKKGDKSEKVVVTKIVER-----ITSEG 864
 QY 416 VDGWYDDATKFTVTTE 433
 Db 865 GDGATKY--ITKSVTVTQ 880

RESULT 5
 EDC8_DAUCA STANDARD; PRT; 555 AA.
 AC P20075;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Embryonic protein DC-8 (Clone 8/10).
 GN DC8.
 OS Daucus carota (Carrot).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
 OX NCBI_TaxID=4039;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. QUEEN ANNE'S LACE;
 RX MEDLINE=8938429; PubMed=2571069;
 RA Franz G., Hatzopoulos P., Jones T.J., Krauss M., Sung Z.R.;
 RT "Molecular and genetic analysis of an embryonic gene, DC 8, from
 Daucus carota L.";
 RL Mol. Gen. Genet. 218:143-151(1989).
 CC -1- FUNCTION: MAY PLAY A ROLE IN LATE EMBRYOGENY.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. PROTEIN BODIES, AND CELL WALLS
 CC OF ZYGOTIC EMBRYO AND ENDOSPERM TISSUE.
 CC -1- SIMILARITY: SOME, TO COTTON LEA D7, TO RAPE LEA 76, AND CARROT
 DC-3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X16131; CAA34258.2;
 DR PIR; S04909; S04909.
 DR InterPro; IPR004238; LEA.
 DR Pfam; PF02987; LEA; 6.
 KW Repeat.
 FT DOMAIN 97 391 17 X APPROXIMATE TANDEM REPEATS.
 FT REPEAT 97 114 1.
 FT REPEAT 115 125 2.
 FT REPEAT 126 140 3.
 FT REPEAT 141 154 4.
 FT REPEAT 155 176 5.
 FT REPEAT 177 191 6.
 FT REPEAT 192 205 7.
 FT REPEAT 206 216 8.
 FT REPEAT 217 237 9.
 FT REPEAT 238 259 10.
 FT REPEAT 260 281 11.
 FT REPEAT 282 303 12.
 FT REPEAT 304 325 13.
 FT REPEAT 326 343 14.
 FT REPEAT 344 358 15.
 FT REPEAT 359 376 16.
 FT REPEAT 377 391 17.
 SQ SEQUENCE 555 AA; 60260 MW; D15B8A30E51BD1AB CRC64;
 Query Match 6.9%; Score 155; DB 1; Length 555;
 Best Local Similarity 23.9%; Pred. No. 0.083;
 Matches 107; Conservative 57; Mismatches 187; Indels 96; Gaps 18;

QY 2 VENKEETPETDSEEVTIKANLIFANGSTOTAEFGKTFEKATSEAYAYADTLKKG 61
 Db 40 VEHEEVSGPG-----VIGSILKSVQGLQAK-----EVVVKAHDTAB-VRENT 86
 QY 62 EYTVADVADGYTLNIFKAGKTEPEEPKEEVTIKANLIYADGKTOTAEFGKTFEATAEA 121
 Db 87 DYAYDKGREGDVAQAQAEAEKAKMAKDTTNGKAGEYKDYTAQKA-----EAEKEA 140
 QY 122 YRYADALKDNGEYTVADVADGYTLNIFKAGKTE-----PEEPKEEVTIKANLIYADGKTQ 178
 Db 141 AQAEETKEAGEY-----KNYTAQKAGEAKDTTLGKAGEYKDYAAQKA-----AEAKDT 190
 QY 179 TAEFGKTFEATAEAAYRYADLLAKNGKYTVADVADGYTLNIFKAGKTEPE-----EPKEEV 236
 Db 191 TAQKAAAEKTEGTE---YKDYAAQKAAEAKVLAQKA-----AEAKDTTGKGEYKDYA 241
 QY 237 TIKANLIYADGK-----TQTAEFGKTFEATAEAAYRYADLLAKNGKYTADLEGGYTTNI 292
 Db 242 AQKA-----AEAKDATMQKTGEYKDYAAQKTAETKDATMEKAEKEYKDYAAQKAAEAKDATM 297
 QY 293 RFAGKKVDEKPEEPMDTYKLIL-NGKTLKGETTTEAVDA--ATAEKV----- 336
 Db 298 QKTGEYKDYSAQKAAETKDATMEKTEYKDYTAQKAAETKDATMEKAEKADTTVQKTGE 357
 QY 337 FKQYANDNGVDGEWTVDDATKFTFTVEKPEVIDASELTPAVTTTKLVINGKTLKGETTK 396
 Db 358 YKDYAAEAKGEG-----KDVTVKAEKGD-----TTVGMKTELKSDAADAARK 401
 QY 397 AVD-----AETAERAKFOY 410
 Db 402 AMDMFLGKKEEVKAGETAEAAKEKY 428

RESULT 6
 MRSP_STAAU STANDARD; PRT; 1637 AA.
 ID MRSP_STAAU
 AC P80544; Q92F62;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Methicillin-resistant surface protein precursor.
 GN PLS.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 OX NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ISOLATE 1061;
 RA Savolainen K., Kuusela P., Paulin L., Korhonen T.K.;
 RT "Pls, a large repeat-rich surface protein of methicillin resistant
 Staphylococcus aureus.";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 624-628; 676-682; 938-948; 1156-1168; 1176-1185;
 RC STRAIN=ISOLATE 1061;
 RX MEDLINE=96270743; PubMed=8665912;
 RA Hilden P., Savolainen K., Tynnelä J., Vuontola M., Kuusela P.;
 RT "Purification and characterisation of a plasmin-sensitive surface
 protein of Staphylococcus aureus.";
 RL Eur. J. Biochem. 236:904-910(1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF115379; AAD09131.1; -.

```

DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR - PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
KW Cell wall; Transmembrane; Antibiotic resistance; Glycoprotein; Repeat;
KW Signal.
FT SIGNAL 1 48 POTENTIAL.
FT CHAIN 49 1637 METHICILLIN-RESISTANT SURFACE PROTEIN.
FT DOMAIN 1301 1582 141 X 2 AA TANDEM REPEATS OF D-[SAG].
FT DOMAIN 1598 1603 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 1637 AA; 174573 MW; 75BE9AD8469BD309 CRC64;

Query Match 6.8%; Score 151.5; DB 1; Length 1637;
Best Local Similarity 24.4%; Pred. No. 0.46;
Matches 111; Conservative 34; Mismatches 200; Indels 109; Gaps 17;

QY 3 ENKEETPETTETSEEEVTTKANLIIFANGSTQTAEPKFGTFEKAATSAAYAYADTLKKDNGE 62
DQ 87 EQVDVTKDITTEQASTEE---KAN-----TTEQASTEERKADTTTEQATTEEAPKAEG- 133
QY 63 YTVADVADKGYTLNKFAGKEKTPPEEKEVTIKANLIYADGKTQTAEFKGTREATAEAY 122
DQ 134 -----TDKVTETEAAPKAETDKAT-----TEEAPKAETDKATEEAP 170
QY 123 RYADALKKDNGETYVDVADKGYTLNKFKA---GKEKTPPEEKEVTIKANLIYA-----D 174
DQ 171 K-----TEETDKATTEEAPAAEETSAAATEEAPKAETSKAATEEAPKAET 217
QY 175 GKQTAEFKGTPEE---ATPEAYRYADLLAKENGKTYVDVADKGYTLNKFAGKEKTPPEE 231
DQ 218 EKTATEEAPKTEETDKVETEAPK-----AEETSAAATEEAPKAETN-----KVETE 267
QY 232 PKEEVTIKANLIYADGKTQTAEPKFGTFAEATA-----EAYRYADLLAKENGKTYADLDG 286
DQ 268 PAEETNKA-----ATETPAVEDTNKASNAOPSETERTQVVDTVAKDLKKSEVTE-- 321
QY 287 GYTINIRFAGKKVDEKPEEPMPTDKYLKILNGKTLKGETTTEAVDAAYAEKVKFYANDNGV 346
DQ 322 -----AEKAEIKVLKPKISNL-----SNEEIKKIALSEVLKETANKENA 361
QY 347 DGEWYTDATKTVTEKPEVIDASELTPAVTYTKLVINGKTLKGETTKKAYVDAETAEK- 405
DQ 362 QPRTAFSVSSNARTNNVYSATLRAAQAODTVTKGTGNTGAHDIIHKTYKEEFPNEG 421
QY 406 ---AFKQYANDN-GVGDGVWYTDAA---TKTFTVT 432
DQ 422 TLTAFNENFNPNGTKGALYNDKIDFNKDFTTIT 455

RESULT 7
SUBF_BACSU STANDARD; PRT; 1433 AA.
AC P16397;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bacillopeptidase F precursor (EC 3.4.21.-) (Esterase) (RP-I protease)
DE (90 kDa serine proteinase).
GN BPR OR BPF.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 195-222.
RX MEDLINE-90170864; PubMed-2106512;
RA Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Theriault K.A.,
RA Pero J.;
RT "Bacillopeptidase F of Bacillus subtilis: purification of the protein
RT and cloning of the gene.";
RL J. Bacteriol. 172:1470-1477(1990).
RN [2]
RP REVISIONS.

```

```

RC STRAIN=168;
RX MEDLINE-90368623; PubMed-2118514;
RA Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Theriault K.A.,
RA Pero J.;
RL J. Bacteriol. 172:5520-5521(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE-90216713; PubMed-2108961;
RA Wu X.-C., Nathoo S., Pang A.S.-H., Carne T., Wang S.-L.;
RT "Cloning, genetic organization, and characterization of a structural
RT gene encoding bacillopeptidase F from Bacillus subtilis.";
RL J. Biol. Chem. 265:6845-6850(1990).
RN [4]
RP SEQUENCE OF 1-211 FROM N.A.
RX MEDLINE-89008108; PubMed-3139638;
RA Beall B., Lowe M., Lutkenhaus J.;
RT "Cloning and characterization of Bacillus subtilis homologs of
RT Escherichia coli cell division genesftsZ andftsA.";
RL J. Bacteriol. 170:4855-4864(1988).
RN [5]
RP SEQUENCE OF 1410-1433 FROM N.A.
RC STRAIN=168 / MARBURG;
RX MEDLINE-90174995; PubMed-2106671;
RA Masuda E.S., Anaguchi H., Sato T., Takeuchi M., Kobayashi Y.;
RT "Nucleotide sequence of the sporulation gene spoIIIGA from Bacillus
RT subtilis.";
RL Nucleic Acids Res. 18:657-657(1990).
RN [6]
RP SEQUENCE OF 195-219.
RC STRAIN=NATTO 16;
RA Kato T., Yamagata Y., Arai T., Ichishima E.;
RT "Purification of a new extracellular 90-kDa serine proteinase with
RT isoelectric point of 3.9 from Bacillus subtilis (natto) and
RT elucidation of its distinct mode of action.";
RL Biosci. Biotechnol. Biochem. 56:1166-1168(1992).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M29035; AAA62679.1; -
CC EMBL; J05400; AAA83362.1; -
CC EMBL; M22630; AAA22458.1; -
CC EMBL; X17344; CAA35224.1; -
CC EMBL; Z99111; CAB13403.1; -
CC EMBL; Z99112; CAB13404.1; -
CC PIR; A35131; A35131.
CC PIR; A35750; A35750.
CC PIR; A36734; A36734.
CC HSP; P00782; 2SPT.
CC MEROPS; S08.017; -.
CC SubtilList; BG10233; bpr.
CC InterPro; IPR000209; Peptidase_S8.
CC Pfam; PF00082; Peptidase_S8; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
CC PROSITE; PS00137; SUBTILASE_HIS; 1.
CC PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolyase; Serine protease; Zymogen; Signal; Complete proteome.
FT SIGNAL 1 30
FT PROPEP 31 194 POTENTIAL.
FT CHAIN 195 755 BACILLOPEPTIDASE F.
FT PROPEP 756 1433 POTENTIAL.
FT ACT_SITE 227 227 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 274 274 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 452 452 CHARGE RELAY SYSTEM (BY SIMILARITY).

```

```
FT CONFLICT 219 219 T -> A (IN REF. 6).
FT CONFLICT 393 393 A -> V (IN REF. 3).
FT CONFLICT 834 834 KHONKA -> N (IN REF. 3).
FT CONFLICT 836 841 QQVLP -> RRLYS (IN REF. 3).
FT CONFLICT 844 852 AQSVVETG -> FCRSRKSV (IN REF. 3).
FT CONFLICT 853 1433 MISSING (IN REF. 3).
SQ SEQUENCE 1433 AA; 154577 MW; 98DF6846897807C9 CRC64;

Query Match 6.7%; Score 149; DB 1; Length 1433;
Best Local Similarity 21.6%; Pred. No. 0.54;
Matches 116; Conservative 59; Mismatches 187; Indels 176; Gaps 27;

QY 2 VENKEETPETPDSSEEVIRKANLIFANGSTQTAEF-KGTFEKATSEAYAYADTLKKDN 60
Db 815 VNPKKAPSANTAVKHONKAIQOVLPKKAQSVVETGKSTYSDQSTGY----TLKHA 870
QY 61 GETVDVAKGYTLNFKAGKEKTPPE-EPKEEYVIRKANLYADGKQTAEFKTFEAT 118
Db 871 GDYTLMAEAYGY------QSTOKVSLKTDQTOANFTLEMKRGT--LKGTVINKTT 920
QY 119 -----AEARYAD-----ALKKDNGEYTVDVADKGYTLNFKAG--KEKTPPEEKVEVT 165
Db 921 GEPVTCASVYVVEDAAVEPAWTDKGEYMLEAYEGATIKVAAPGYSDSEFVELKGDVT 980
QY 166 -----IKANLYADGKTQTA----- 180
Db 981 KETALPKFVGYPGEIAYDGTAEANANSYFAAGNGWAKVMTLADGDKGMLTGLFRFWDT 1040
QY 181 -----EPKGFTEAT-----AEARYADLLAKENGKKT-VDVADKG- 215
Db 1041 EFPDPGTEFKEVYDATGDKGAPRKIAKGFPAEALR-----NGETKVDLSKGI 1092
QY 216 -----YTLNFKAGKEKTPPEEKVEVTKA--NLVYADGKQTQ-----TAEF 254
Db 1093 MVDKDFLYVIQSKPDPSPLGANDGQNGRNWQIDGKWPGKADGNVIMRALVDY 1152
QY 255 KGTFEATA---BARYAD--LLAKENGKYTDLEDGGYTNIRFAGKVDKPEEPMDTY 310
Db 1153 EAAVPEITSPDKSYNTKDSVTYKQNA-----SPGTTVHIYNGEKEAGETKAAADGTF 1205
QY 311 K--LIINGTLKGETTEAVDAATAEKVQYANDNGVDGEWYDDATKTFVT---EKP 365
Db 1206 HAGIILN---KGENEL---TATA-----STDNGT-----TDASSPITVTLDOEKP 1244
QY 366 EVI-----DASELTPAVTVYKLVINGTKL-----GETTKAVDAETAERAKFOYAND 413
Db 1245 ELTLDNPKDGGKNTKETLTVKGAVSDNLDKVKVNGKKATVADGYSARILLENGNE 1302

RESULT 8
ABP2_HUMAN STANDARD; PRT; 2647 AA.
AC P21333;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endothelial actin-binding protein (ABP-280) (Nonmuscle filamin)
DE (Filamin 1) (Alpha-filamin).
GN FLN1 OR FLN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90361737; PubMed=2391361;
RA Gorlin J.B., Yamin R., Egan S., Stewart M., Stossel T.P.,
RA Kwiatkowski D.J., Hartwig J.H.;
RT "human endothelial actin-binding protein (ABP-280, nonmuscle
RT filamin): a molecular leaf spring.";
RL J. Cell Biol. 111:1089-1105(1990).
RN [2]
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE=96311563; PubMed=8733135;
RA Chen E.Y., Zollo M., Mazzarella R.A., Ciccodicola A., Chen C.-N.,
RA Zuo L., Heiner C., Burrough F.W., Ripetto M., Schlessinger D.,
RA D'Urso M.;
RT "Long-range sequence analysis in Xq28: thirteen known and six
RT candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and
RT G6PD loci.";
RL Hum. Mol. Genet. 5:659-668(1996).
RN [3]
RP SEQUENCE OF 1658-1772 FROM N.A.
RX MEDLINE=93357748; PubMed=7689010;
RA Maestrini E., Patrosso C., Mancini M., Rivella S., Rocchi M.,
RA Repetto M., Villa A., Frattini A., Zoppe M., Vezzoni P.,
RA Toniolo D.;
RT "Mapping of two genes encoding isoforms of the actin binding protein
RT ABP-280, a dystrophin like protein, to Xq28 and to chromosome 7.";
RL Hum. Mol. Genet. 2:761-766(1993).
CC -!- FUNCTION: PROMOTES ORTHOGONAL BRANCHING OF ACTIN FILAMENTS AND
CC LINKS ACTIN FILAMENTS TO MEMBRANE GLYCOPROTEINS.
CC -!- SUBUNIT: HOMODIMER. INTERACTS WITH CVHSP.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL CYTOPLASM.
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -!- PTM: PHOSPHORYLATION EXTENT CHANGES IN RESPONSE TO CELL
CC ACTIVATION.
CC -!- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY
CC ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,
CC ABP-120, ABP-180, OR BETA-FODRIN).
CC -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -!- SIMILARITY: CONTAINS 24 FILAMIN REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X53416; CAA37495.1; -
CC EMBL; L44140; AAA2644.1; -
CC EMBL; X70082; CAA49687.1; -
CC EMBL; X70085; CAA49690.1; -
CC PIR; A37098; A37098.
CC HSSP; P13466; 1KSR.
CC MIM; 300017; -
CC InterPro: IPR001589; Actinin_act_bind.
CC InterPro: IPR001715; Calponin_hom.
CC InterPro: IPR001298; Filamin.
CC Pfam; PF00307; CH; 2.
CC Pfam; PF00630; Filamin; 24.
CC SMART; SM00033; CH; 2.
CC PROSITE; PS00019; ACTININ_1; 1.
CC PROSITE; PS00020; ACTININ_2; 1.
CC PROSITE; PS00021; CH; 2.
CC PROSITE; PS0194; FILAMIN_REPEAT; 24.
CC Actin-binding; phosphorylation; Repeat.
FT DOMAIN 1 274 ACTIN-BINDING (HEAD).
FT DOMAIN 43 149 CH 1.
FT REPEAT 166 266 CH 2.
FT REPEAT 276 374 FILAMIN 1.
FT REPEAT 376 474 FILAMIN 2.
FT REPEAT 475 570 FILAMIN 3.
FT REPEAT 571 663 FILAMIN 4.
FT REPEAT 667 763 FILAMIN 5.
FT REPEAT 764 866 FILAMIN 6.
FT REPEAT 867 965 FILAMIN 7.
FT REPEAT 966 1061 FILAMIN 8.
FT REPEAT 1062 1154 FILAMIN 9.
FT REPEAT 1155 1249 FILAMIN 10.
FT REPEAT 1250 1349 FILAMIN 11.
FT REPEAT 1350 1442 FILAMIN 12.
FT REPEAT 1443 1539 FILAMIN 13.
FT REPEAT 1540 1636 FILAMIN 14.
```

```
FT REPEAT 1649 1740 FILAMIN 15.
FT DOMAIN 1741 1778 HINGE 1.
FT REPEAT 1779 1860 FILAMIN 16.
FT REPEAT 1861 1950 FILAMIN 17.
FT REPEAT 1951 2039 FILAMIN 18.
FT REPEAT 2042 2131 FILAMIN 19.
FT REPEAT 2132 2230 FILAMIN 20.
FT REPEAT 2233 2325 FILAMIN 21.
FT REPEAT 2327 2420 FILAMIN 22.
FT REPEAT 2424 2516 FILAMIN 23.
FT DOMAIN 2517 2551 HINGE 2.
FT REPEAT 2552 2646 FILAMIN 24.
FT DOMAIN 2557 2647 SELF-ASSOCIATION SITE, TAIL.
FT SITE 1761 1762 CLEAVAGE (BY CALPAIN).
FT MOD_RES 1 1 BLOCKED.
FT CONFLICT 1772 1772 A -> G (IN REF. 3).
FT CONFLICT 2634 2634 H -> D (IN REF. 2).
SQ SEQUENCE 2647 AA; 280759 MW; 6C1A07041DFA3D42 CRC64;

Query Match 6.6%; Score 148.5; DB 1; Length 2647;
Best Local Similarity 21.8%; Pred. No. 1.2;
Matches 116; Conservative 66; Mismatches 190; Indels 159; Gaps 28;

QY 47 SEAYAYADILKNDNGYTVVDVADKG---YTLNFKFAGKEKTEPEP---KEEVYIRANLIY 100
Db 818 AEADIDFDIIRNDNDFTVKTYPRGAGSYTIMVLEA-DQATPTSPTRVKEVPSHDASKVK 876
QY 101 ADCK-----TQTAFFKG-----TFEAT-AEAYRYADALKKNDNGEYTV 138
Db 877 AEGPGLSRGTGVELGKTHFTVNAKAAGKGLDVQFSLGTLKGAQVAVRDVIDHNDNTYTVK 936
QY 139 VADKGYT-----LNIRFAGKEKTEPEP-----KEE 163
Db 937 -----YTPVQGGPGVGNVYGG-DPIPKSPFSAVSPSLDKIKVSLGEXVDVCKDOE 990
QY 164 VTIKANLIYADKG--TQTAFFKGTFPEATAEAYRYAD-----LLAKENGYTVVDVADKGY 216
Db 991 FTVKSGAGGOGKVASKIVGSPGAAPVCPVEPLGADNSVVRFLPREEGPYEVEVYDGV 1050
QY 217 TL-NIRFAGKEKTEPEPKEEVTIKANLIYADG-----KTQTAFFKGTFPEATAE 263
Db 1051 PVPGPSFPLEAVAPTRPSK---VKA---FGPGLQGSAGSPARFTIDTKAGTGGGLGLTV 1104
QY 264 EAYRYADLLAKENGYTADL-----EDGGYTTINIRFAGKKVDEKPEP-----MDTYKLI 313
Db 1105 EGPCEAQLECLDNGDGTCSVSVYPTPEPDYINILFADTHIPGSPKHAHVPCFDASKVK 1164
QY 314 LNKGLTKGETTTEA---VDAATA--EKVKQYANDNGVDGE---WTYDQATKT-----358
Db 1165 CSGPGLERATAGEVGQFQVDCSAGSAELTIEICSEAGLPAEYVIQDHGDGTHITYIPL 1224
QY 359 ----FTVTEK-----PEVIDASELTPAVTYTKLVINGKTLKE-----TTTKAVDAET 402
Db 1225 CPQAYVTYIKYGGQVPNPFPSKLQVPAVDTSVQCYGFCIGQGQVFRATTEFSVDARA 1284
QY 403 AEKA-----FKQYANDNGVDGVW-----TYDDATKTFVT 432
Db 1285 LTQTGGPHVKARVANPSGNLITYVQDRG-DGMVKVEYTPYEGLHSVDVT 1334

RESULT 9
YF16_YEAST STANDARD; PRT; 1233 AA.
AC P43597;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 137.7 kDa protein in UGSI-FAB1 intergenic region.
GN YFR016C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
```

```

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=S288C / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae."
RL Nat. Genet. 10:261-268(1995).
CC -!- SIMILARITY: SOME, TO MAMMALIAN NEUROFILAMENT TRIPLET M PROTEIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D50617; BAA09255.1; -
DR SGD: S0001912; YFR016C.
KW Hypothetical protein.
SQ SEQUENCE 1233 AA; 137697 MW; C8A7CD2C6F0892F6 CRC64;

Query Match 6.6%; Score 148; DB 1; Length 1233;
Best Local Similarity 21.2%; Pred. No. 0.51;
Matches 99; Conservative 64; Mismatches 181; Indels 124; Gaps 16;

QY 2 VENKSETETPTDSEEVYTIKANLIIFANGSTQTAFFKGTPEKATS--EAYAYADTLKDD 59
Db 465 VEKEEEEEEEENSTFSKVKKENVTGQEAVRNNEVSGTEESTSKGEEIMGGDEKQSE 524
QY 60 NGEYTVVDVADKGYTLNIRFAGKEKTEPEPKEEVTIKANLIYAD-GKTQTAFFKGT---- 113
Db 525 AGEKSIIEIGSANSKIS-----KDNLVLEDEAEAPTQENKPTVEVGE 569
QY 114 --FEATAEAYRYADALKKNDNGEYTVVDVADKGYTLNIRFAGKEKTEPEPKEEVTIKANLI 171
Db 570 IDIPADPRDDVEIVEAVEKNIIPEDLEVAKEDQ-----EGQVKLDEPVK--AMKDDKI 621
QY 172 YADGKTQTAFFKGTPEATAEAYRYADLLAKENGYTVVDVADKGYTLNIRFAGKEKTEPEE 231
Db 622 AMRGAESISEDMKKQEGTAE-----LSNEKAKKEVDETARESAEGVE--VEKSKTPES 673
QY 232 PK-----EVTIKANLIYADCKTQTAFFK 255
Db 674 PKVVRCTSGRPEDLIINERDPEVLKEDVVRPEDEKVPKEIATTIENSEEDPKSORVQIS 733
QY 256 GTFAEATAEAYRYADLLAKENGYTADLEGGYTTINIRFAGKKVDEKPEPMDTYKLIIN 315
Db 734 TEQAETTKQ-----DMGDVGSTTSF-----KEEEKPKR---FEITQE 767
QY 316 GYTLKGETT-----TEAVDAATAEAKVF-----KQYANDNGVDGWTYDDATK 357
Db 768 GDKITGKDTNHHGEATEAASNSKASDVGTAIEPSSSVKDKDTEADAENVENSEKTE 827
QY 358 TPTVTEKPEVIDASELTPAVTYTKLVINGKTLKGETTTTKAVDAETAEK 405
Db 828 FIKVKAELNLDAPK--EAEVTAELNKENEDVEVDTEEDA--EVENSEK 872

RESULT 10
DXT_STRDO STANDARD; PRT; 1337 AA.
AC P39653;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE Dextranase precursor (EC 3.2.1.11) (Alpha-1,6-glucan-6-
DE glucanohydrolase).
OS Streptococcus downei (Streptococcus sobrinus).
```


OG Plasmid pYA902.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1317;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-36.
 RC STRAIN=6715 / UAB66;
 RX MEDLINE=94292401; PubMed=80211165;
 RA Wanda S.-Y., Curtiss R. III;
 RT "Purification and characterization of Streptococcus sobrinus
 RT dextranase produced in recombinant Escherichia coli and sequence
 RT analysis of the dextranase gene."
 RL J. Bacteriol. 176:3839-3850(1994).
 CC -!- FUNCTION: MAY PLAY A ROLE IN SUCROSE-INDEPENDENT ADHERENCE TO THE
 CC PELICLE-COATED TOOTH SURFACE. THE ACTIVITY OF THIS ENZYME IS
 CC OPTIMAL AT PH 5.3 AND AT 39 DEGREES CELSIUS.
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,6-alpha-D-glucosidic
 CC linkages in dextran.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
 CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
 CC IN THE REGION OF THE MEMBRANE ANCHOR.
 CC -!- SIMILARITY: BELONGS TO FAMILY 66 OF GLYCOSYL HYDROLASES.
 CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-5 IS THE INITIATOR.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M96978; AAA21772.1;
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
 KW Hydrolase; Glycosidase; Signal; Transmembrane; Repeat; Plasmid.
 FT SIGNAL 1 30
 FT CHAIN 31 1337 DEXTRANASE.
 FT DOMAIN 31 1313 EXTRACELLULAR (BY SIMILARITY).
 FT TRANSMEM 1314 1332 MEMBRANE ANCHOR (BY SIMILARITY).
 FT DOMAIN 1333 1337 CYTOPLASMIC (BY SIMILARITY).
 FT DOMAIN 1305 1310 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
 FT PROTEINS.
 SQ SEQUENCE 1337 AA; 143298 MW; B494275A77A2E3D0 CRC64;
 Query Match 6.6%; Score 146.5; DB 1; Length 1337;
 Best Local Similarity 20.9%; Pred. No. 0.68;
 Matches 110; Conservative 64; Mismatches 181; Indels 171; Gaps 26;
 QY 6 EETPETPET-DSEEEVTIKANLI--FANGSTQ----TAEFGTFEKATSEAYAYADTLKK 58
 DB 100 EQTP--PETSASAPATTSADSVKAYQADATQNSTSGNGPVIRATSAQVTRATRSVSS 157
 QY 59 DNGEYTDV-ADK-----GYTLNIFAGKEKTPPEEKVEVTIKANLIYADGKTQTAEFRG 112
 DB 158 QSGDAIVLSADKASRYQEDVNLSDFKNTTDE--QDVTVDVYIIDNKLGT----- 210
 QY 113 TEEATAEAYRADALKKNGEYTDVADKGYTLNIFAGKEKTPP---EPKEEVTIKAN 169
 DB 211 -----YKFSKHLK-----AGEGYKMQ---SGDLKIPASQFENNHHYLLKVR 248
 QY 170 LIYADGKTQTAEFGTFEKATEAEAY-RYA-----DLLAKENGKVTVDVAD-KGYT 217
 DB 249 VRDADNLTUSEVKAIAVESDWTKPPRYGIVGSGQDNTNLSLKADRYAEIEKMKNN 308
 QY 218 LNIKF---AGKEKTPPEEKVEVTIKA-----NL 242
 DB 309 INSFFDYDYKTKATNPFPSDEATFKQDWNWSSSEIDTQAVKQIVNQVHGGAVAMLYNM 368
 QY 243 IYADGKTQTAEF-----KGTFP-----EATAEAYRVDLL 272

DB 369 ILAE-NTNTGEAPVLPETEYAYNSDDRGYGAQGPMSTYTVKIPKGDGEDEVEIQRYNPT 427
 QY 273 AKENGKYTAD-----LEDGGY-----TINIFAGKKVDEKPEEPMDTYKLIILNGTKLG 321
 DB 428 SKLWQDYIADKMGQAMKNGGFGDQWQDITGDNEVYSYADKDSNDPSKKFWL----- 478
 QY 322 ETTTEAVD---AATAEKVFKFYANDNGVDGE-----WTYDDATKTFV 361
 DB 479 ---TEGYAEFLRAIKELPNLYLTVDNNGEQIYRLKDGQNDVIYNEINPFGPALPSEMA 535
 QY 362 TEKPEVIDASELTPAVT--TYKLVIINGKTLKG-ETTTKAVDAETA 404
 DB 536 AVKPNTVTSRPVLTKVRQGDWKISIVGAYMEGSENGSKADAERAGK 581
 RESULT 11
 WAPA_BACSU STANDARD; PRT: 2334 AA.
 ID Q07833;
 AC 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Wall-associated protein precursor.
 GN WAPA OR NI7G.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=93302506; PubMed=8316082;
 RA Foster S. J.;
 RT "Molecular analysis of three major wall-associated proteins of
 RT Bacillus subtilis 168: evidence for processing of the product of a
 RT gene encoding a 258 kDa precursor two-domain ligand-binding
 RT protein."
 RL Mol. Microbiol. 8:299-310(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSC1A1;
 RX MEDLINE=95219088; PubMed=7704263;
 RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
 RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
 RT genome containing the hut and wapa loci."
 RL Microbiology 141:337-343(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSC1A1;
 RX MEDLINE=97124196; PubMed=8969509;
 RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
 RA Miwa Y., Fujita Y.;
 RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
 RT containing the lic and cel loci, and creation of a 177 kb contig
 RT covering the gnt-sacxy region."
 RL Microbiology 142:3113-3123(1996).
 CC -!- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
 CC MOTILITY, SECRETION OR DIFFERENTIATION.
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
 CC INTO THE MEDIUM.
 CC -!- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
 CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
 CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
 CC MOTIF REPEATED 31 TIMES.
 CC -!- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
 CC SIMILARITY TO THE REPEAT IN E. COLI RHS GROUP OF PROTEINS (RHS-A-D).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC

```
CC or send an email to license@sib.ch).
CC -----
DR EMBL; L05634; AAA22883.1; -
DR EMBL; D31856; BAA06656.1; -
DR EMBL; D29985; BAA06260.1; -
DR EMBL; D83026; BAA11683.1; -
DR EMBL; Z99124; CAB15959.1; -
DR PIR; S32920; S32920.
DR Subtilist; BG10797; wapA.
DR InterPro; IPR003305; CBD_6.
DR Pfam; PF02018; CBD_6; 1.
KW SIGNAL 1 28 Complete proteome.
FT CHAIN 29 2334 OR 32 (POTENTIAL).
FT DOMAIN 504 869 WALL-ASSOCIATED PROTEIN.
FT REPEAT 504 605 3 X 101 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 636 736 1-1.
FT REPEAT 769 869 1-2.
FT DOMAIN 1021 2139 1-3.
FT REPEAT 1021 2139 31 X 21 AA APPROXIMATE TANDEM REPEATS OF
FT X(4)-G-X(4)-[YF]-X-D-X(2)-G-X(4).
FT REPEAT 1021 1040 2-1.
FT REPEAT 1042 1061 2-2.
FT REPEAT 1063 1082 2-3.
FT REPEAT 1083 1102 2-4.
FT REPEAT 1109 1128 2-5.
FT REPEAT 1129 1148 2-6.
FT REPEAT 1150 1169 2-7.
FT REPEAT 1174 1193 2-8.
FT REPEAT 1199 1218 2-9.
FT REPEAT 1219 1238 2-10.
FT REPEAT 1646 1665 2-11.
FT REPEAT 1667 1686 2-12.
FT REPEAT 1690 1709 2-13.
FT REPEAT 1711 1730 2-14.
FT REPEAT 1732 1751 2-15.
FT REPEAT 1753 1772 2-16.
FT REPEAT 1795 1814 2-17.
FT REPEAT 1820 1839 2-18.
FT REPEAT 1840 1859 2-19.
FT REPEAT 1861 1880 2-20.
FT REPEAT 1887 1906 2-21.
FT REPEAT 1908 1927 2-22.
FT REPEAT 1929 1948 2-23.
FT REPEAT 1969 1982 2-24.
FT REPEAT 1983 2002 2-25.
FT REPEAT 2008 2027 2-26.
FT REPEAT 2028 2047 2-27.
FT REPEAT 2051 2070 2-28.
FT REPEAT 2071 2090 2-29.
FT REPEAT 2093 2112 2-30.
FT REPEAT 2120 2139 2-31.
SQ SEQUENCE 2334 AA; 258329 MW; B75138CCDD278BAA3 CRC64;

Query Match 6.4%; Score 144; DB 1; Length 2334;
Best Local Similarity 21.1%; Pred. No. 1.8;
Matches 118; Conservative 70; Mismatches 214; Indels 156; Gaps 27;

QY 14 TDSEEVTTKANLIFANGSTQTAERKGTTEKATSEAYAVADTLKKDNGEYTVDA---DK 70
DB 1534 TD-EKDYSLQANVTYADGS--TGIYNAKTPSGTQENRAAVIPKTPINKVDISILFQK 1590
QY 71 GYTLNFKF-----AGKEKTPPEEKVTTKANLIYADGKTOTAEFKG 112
DB 1591 SARGTWVFDDIRLEGSLLTKSTYDSNGNVTYKEEDLGYATSTDYDEGKKTSETDAKG 1650
QY 113 TFEATAEAYRYADAL-----KKDNGEYTVDV---ADKGYTLNFKPAGK 154
DB 1651 ---EKTYYDQDLQTLNLSNGTSILHSYDKGEGNEVSKTRAGADQTYKFEYDVMGKL 1707
QY 155 KTPPEEP-----KEVTTKANL---IYADGKTOTAEFKGTFEEATAPAY---RYADLLAK 202
DB 1708 VKTTDPLGNVLASEYDANSLTKTISPNGNEVSLSDYGT-DRVKSXSGNGTEKYIFTYDK 1766
```


AC P15205; Q62958; Q9ER21; Q9QW92;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Microtubule-associated protein 1B (MAP1B) (Neuraxin) [Contains: MAP1
 DE light chain LCL].
 GN MAP1B.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE OF 1-142 FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Testis;
 RX MEDLINE=96257242; PubMed=8666295;
 RA Liu D., Fischer I.;
 RT "Isolation and sequencing of the 5' end of the rat microtubule-
 RT associated protein (MAP1B)-encoding cDNA.";
 RL Gene 171:307-308(1996).
 RN [2]
 RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain, and Glial tumor;
 RX MEDLINE=92347374; PubMed=1639092;
 RA Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;
 RT "Identification of two distinct microtubule binding domains on
 RT recombinant rat MAP 1B.";
 RL Eur. J. Cell Biol. 57:66-74(1992).
 RN [3]
 RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
 RC TISSUE=Spinal cord;
 RX MEDLINE=90059871; PubMed=2555150;
 RA Rientz A., Grenningloh G., Hermans-Borgmeyer I., Kirsch J.,
 RA Littauer U.z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
 RT "Neuraxin, a novel putative structural protein of the rat central
 RT nervous system that is immunologically related to microtubule-
 RT associated protein 5.";
 RL EMBO J. 8:2879-2888(1989).
 RN [4]
 RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
 RX MEDLINE=97405699; PubMed=9260743;
 RA Ma D., Nothias F., Boyne L.J., Fischer I.;
 RT "Differential regulation of microtubule-associated protein 1B (MAP1B)
 RT in rat CNS and PNS during development.";
 RL J. Neurosci. Res. 49:319-332(1997).
 CC -!- FUNCTION: The function of brain MAPS is essentially unknown.
 CC Phosphorylated MAP1B may play a role in the cytoskeletal changes
 CC that accompany neurite extension. Possibly MAP1B binds to at least
 CC two tubulin subunits in the polymer, and this bridging of subunits
 CC might be involved in nucleating microtubule polymerization and in
 CC stabilizing microtubules.
 CC -!- SUBUNIT: 3 different light chains, LCL1, LCL2 and LCL3, can associate
 CC with MAP1A and MAP1B proteins.
 CC -!- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
 CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
 CC heart or muscle.
 CC -!- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
 CC nerve levels are high early in development but decrease during
 CC postnatal development and are low in adults. In dorsal root
 CC ganglia levels remain high throughout development.
 CC -!- INDUCTION: By nerve growth factor.
 CC -!- DOMAIN: Has a highly basic region with many copies of the sequence
 CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
 CC responsible for the binding of MAP1B to microtubules.
 CC -!- PTM: LCL1 is coexpressed with MAP1B. It is a polypeptide generated
 CC from MAP1B by proteolytic processing. It is free to associate with
 CC both MAP1A and MAP1B. It interacts with the amino-terminal region
 CC of MAP1B (by similarity).
 CC -!- PTM: Phosphorylated.
 CC -!- SIMILARITY: TO MAP1A.
 CC -!- CAUTION: A C-terminal fragment of this protein (residues 1597 to
 CC 2459) was originally described as neuraxin in ref.3.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U52950; AAB17068.1; -
 DR EMBL; X60370; CAC16162.1; -
 DR EMBL; X16623; CAA34620.1; ALT_SEQ.
 DR PIR; S06017; S06017.
 DR InterPro; IPRO00102; MAP1B_neuraxin.
 DR Pfam; PF00414; MAP1B_neuraxin; 10.
 DR PROSITE; PS00230; MAP1B_NEURAXIN; 8.
 KW Microtubules; Repeat; Phosphorylation.
 FT CHAIN ? 2459 MAP1 LIGHT CHAIN LCL.
 FT REPEAT 1869 1885 MAP1B 1.
 FT REPEAT 1886 1902 MAP1B 2.
 FT REPEAT 1903 1919 MAP1B 3.
 FT REPEAT 1920 1936 MAP1B 4.
 FT REPEAT 1937 1953 MAP1B 5.
 FT REPEAT 1954 1970 MAP1B 6.
 FT REPEAT 1988 2004 MAP1B 7.
 FT REPEAT 2005 2021 MAP1B 8.
 FT REPEAT 2022 2038 MAP1B 9.
 FT REPEAT 2039 2055 MAP1B 10.
 FT DOMAIN 559 1035 GLU-RICH.
 FT DOMAIN 588 786 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
 FT KKEE AND KKEI/V REPEATS).
 FT DOMAIN 2224 2312 LYS-RICH.
 FT CONFLICT 127 127 M -> V (IN REF. 1).
 FT CONFLICT 140 140 T -> S (IN REF. 1).
 FT CONFLICT 2112 2112 R -> K (IN REF. 3).
 FT CONFLICT 2169 2169 L -> I (IN REF. 3).
 SQ SEQUENCE 2459 AA; 269497 MW; 2E3F6872DEDB8BA2 CRC64;
 Query Match 6.3%; Score 141; DB 1; Length 2459;
 Best Local Similarity 23.4%; Pred. No. 2.8;
 Matches 96; Conservative 59; Mismatches 159; Indels 96; Gaps 22;
 QY 6 EETPEPTDSEEVIVKANLIKANGSTQTAEPKGFPEKATSEAYAYADTLK-----KDN 60
 DB 1698 EEPSTYQDNLSLISV--SQVEASPTSSAHTPSQIASPLQE-----DTLSDVVPDRM 1750
 QY 61 GEYTVADVADKYLNIKFKAGKEKTPPEKPEEVNIK-ANLIYADG-----KTQAEFK 111
 DB 1751 SLVSLASEKVSQLE---CEKLSPKSDISPLTPRESSPTSPGFSDSSTGAKESTAAVQ 1806
 QY 112 GT----FEEATAIRYAD-----ALKKNGEYTVADVADKGYT---LNKIFAGK 153
 DB 1807 TSSSPPIDAAAAPYGRSSMLFDTMQHHLASRLDLTSSVEKDNKGKTPGDFENYAYOKP 1866
 QY 154 EKTPEEPKE-----EVIKANLI--YADGKTQTAEPKGFPEEATAEAYRADLLAKE 203
 DB 1867 ESTSPDEEDYDYESHEKTIQAHVGGVYVEKTER-----TIKSPDSCSYSETI---- 1917
 QY 204 NGYTVADVADKGYTLNKFAGKEKTPPEKPE-----EVIKANLIYADGKTQTAEPKGF 258
 DB 1918 -EKTTPEDGGYSCSEI-----TEKTRTPPEGYSYSEISEKT-----TRTPESGYT 1964
 QY 259 AEATAEAYRADLLAKENGKYTADLEDGGYTIINIRFAGKRVDEK-----PEEPMDTYKLL 314
 DB 1965 YEKTERSRLLDDIS--NG--YDDEGGHGLGDCSYSYETTEKITSFPES--ESYSYET 2018
 QY 315 NGKTLGETTTEAVDAATAEKVKQVANDNGVDGEWYTDATKTFVTVEK 364
 DB 2019 TTKTRSPOTS-AYCVETMEKITK-----TPQASTYSYETSRCYTPERK 2062
 Search completed: October 29, 2002, 09:30:03
 Job time : 19.571 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:27:36 ; Search time 18.2057 Seconds
(without alignments)
2290.646 Million cell updates/sec

Title: US-08-325-278b-3
Perfect score: 2235
Sequence: 1 AVENKEETPETDSEEV.....GVDGWTYDDATKFTVTM 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2235	100.0	455	2 A45063	immunoglobulin-bin
2	1587.5	71.0	719	2 A42808	Ig light chain-bin
3	1263.5	56.5	992	2 S54396	protein L precursor
4	708	31.7	593	2 S00128	protein G precursor
5	706.5	31.6	664	2 S42574	streptococcal surf
6	683	30.6	448	2 A24496	IgG-binding protei
7	404	18.1	413	2 S55890	plasma protein rec
8	188.5	8.4	1185	2 A42404	collagen adhesin -
9	182	8.1	479	2 T47561	late embryogenesis
10	175	7.8	448	2 H84782	late embryogenesis
11	171.5	7.7	74	2 A34483	Ig light chain-bin
12	168	7.5	448	2 JC6171	late embryogenesis
13	162	7.2	913	2 TS2485	neurofilament prot
14	159	7.1	1856	2 C95008	immunoglobulin A1
15	159	7.1	2551	2 B98047	hypothetical prote
16	157.5	7.0	1166	2 T28680	fibrinogen-binding
17	157	7.0	798	2 I50479	neurofilament medi
18	156.5	7.0	701	2 H98120	choline binding pr
19	156	7.0	916	2 A27864	neurofilament trip
20	155	6.9	555	2 S04909	embryonic protein
21	154.5	6.9	1110	2 I51116	NF-180 - sea lamp
22	154.5	6.9	1822	2 S33441	EF protein - Strep
23	154	6.9	737	2 H81070	lactoferrin-bindin
24	154	6.9	1612	2 AB1347	probable peptidogl
25	153.5	6.9	1873	2 T30944	surface protein pr
26	152	6.8	3147	2 T18674	hypothetical prote
27	150	6.7	1489	2 T31108	cyst germination s
28	150	6.7	3488	2 T34418	hypothetical prote
29	149.5	6.7	744	2 F95013	pneumococcal surfa

ALIGNMENTS

RESULT 1

A45063
immunoglobulin-binding protein LG - Peptostreptococcus magnus
C:Species: Peptostreptococcus magnus
C:Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
C:Accession: A45063
R:Kihlberg, B.M.; Sjobring, U.; Kastern, W.; Bjorck, L.
J. Biol. Chem. 267, 25583-25588, 1992
A:Title: Protein LG: a hybrid molecule with unique immunoglobulin binding properties.
A:Reference number: A45063; MUID:93094283
A:Accession: A45063
A>Status: preliminary
A:Molecule type: mRNA; protein
A:Residues: 1-455 <KIH>
A:Cross-references: GB:S50809; NID:g261705; PIDN:AAA03280.1; PID:g261706
A>Note: sequence extracted from NCBI backbone (NCBIN:120302, NCBIP:120303)

Query Match 100.0%; Score 2235; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.9e-110;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AVENKEETPETDSEEV	IKANLIFANGSTQTA	AEFKGTAEFKGTAE	YAYADTLTKDN	60
DB	22	AVENKEETPETDSEEV	IKANLIFANGSTQTA	AEFKGTAEFKGTAE	YAYADTLTKDN	81
QY	61	GEYTVADVADKGYTLN	IKFAGKEKTPPEEK	BEVTIKANLIYADGKT	QTAEFKGTFEEATAE	120
DB	82	GEYTVADVADKGYTLN	IKFAGKEKTPPEEK	BEVTIKANLIYADGKT	QTAEFKGTFEEATAE	141
QY	121	AYRYADALKKDNGEY	TVADVADKGYTLN	IKFAGKEKTPPEEK	BEVTIKANLIYADGKT	180
DB	142	AYRYADALKKDNGEY	TVADVADKGYTLN	IKFAGKEKTPPEEK	BEVTIKANLIYADGKT	201
QY	181	ERKGTPEETAETAE	RYADLLAKENKGYT	VDVADKGYTLN	IKFAGKEKTPPEEK	240
DB	202	ERKGTPEETAETAE	RYADLLAKENKGYT	VDVADKGYTLN	IKFAGKEKTPPEEK	261
QY	241	NLIYADGKTQTAE	FKGTAEATAEAY	RYADLLAKENKGYT	ADLEGGYTTINIRFAG	300
DB	262	NLIYADGKTQTAE	FKGTAEATAEAY	RYADLLAKENKGYT	ADLEGGYTTINIRFAG	321
QY	301	EKPEEPMDYTKLL	NGKTLKGETT	TEAVDAATAEK	VFKQYANDNGVDG	360
DB	322	EKPEEPMDYTKLL	NGKTLKGETT	TEAVDAATAEK	VFKQYANDNGVDG	381
QY	361	VTEKPEVIDASE	LTPAVTVTKLVN	CKTLKGETTT	TKAVDAETAETAK	420
DB	382	VTEKPEVIDASE	LTPAVTVTKLVN	CKTLKGETTT	TKAVDAETAETAK	441
QY	421	TYDDATKTTFT	VTTEM	434		
DB	442	TYDDATKTTFT	VTTEM	455		

30	149.5	6.7	938	2	AFI772
31	149	6.7	1433	1	A36734
32	149	6.7	1876	2	E97944
33	148.5	6.6	990	2	H88733
34	148.5	6.6	2647	2	A37098
35	148	6.6	1233	2	S56271
36	148	6.6	2508	2	S61441
37	147.5	6.6	657	2	AD1525
38	147.5	6.6	1099	2	T14850
39	147.5	6.6	1229	2	T25697
40	147.5	6.6	6642	2	T29757
41	147	6.6	745	2	B33856
42	147	6.6	891	2	G41662
43	146.5	6.6	1337	2	T30291
44	146.5	6.6	1363	2	C84346
45	146.5	6.6	1993	2	AFI450

internalin-like pr
bacillopeptidase F
zinc metalloprotei
protein F32E10.3
gelation factor AB
hypothetical prote
surface-associated
probable cell surf
S-layer protein pr
hypothetical prote
protein UNC-89 - C
130K surface exclu
dextranase - Strep
hypothetical prote
probable peptidogl

RESULT 2

A42808
Ig light chain-binding protein precursor - Peptostreptococcus magnus
N:Alternate names: protein L
C:Species: Peptostreptococcus magnus
C:Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 07-May-1999
C:Accession: A42808; A41493
R:Kastern, W.; Sjoerbring, U.; Bjoerck, L.
J. Biol. Chem. 267, 12820-12825, 1992, 1990
A:title: Structure of peptostreptococcal protein L and identification of a repeated immunoglobulin domain
A:Reference number: A42808; MUID:92316971
A:Accession: A42808
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-719 <KAS>
A:Cross-references: GB:M86697
R:Kastern, W.; Holst, E.; Nielsen, E.; Sjoerbring, U.; Bjoerck, L.
Infect. Immun. 58, 1217-1222, 1990
A:title: Protein L, a bacterial immunoglobulin-binding protein and possible virulence determinant
A:Reference number: A41493; MUID:90215984
A:Accession: A41493
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 202-275 <KA2>
C:Keywords: immunoglobulin

Query Match 71.0%; Score 1587.5; DB 2; Length 719;
Best Local Similarity 76.3%; Pred. No. 3.2e-76;
Matches 334; Conservative 13; Mismatches 56; Indels 35; Gaps 6;

QY 2 VENKEETPETPDSEEEVTKANLIFANGSTQTAFFKGTFFKATSEAYAYADTLKKDNG 61
DB 95 VENKEETPETPDSEEEVTKANLIFANGSTQTAFFKGTFFKATSEAYAYADTLKKDNG 194

QY 62 EYTVADVADKGYTLNIFAGKEKTPPEPKKEVTKANLIYADGKTQTAFFKGTFFKATSEAYAYADTLKKDNG 121
DB 155 EYTVADVADKGYTLNIFAGKEKTPPEPKKEVTKANLIYADGKTQTAFFKGTFFKATSEAYAYADTLKKDNG 214

QY 122 YRVADALKKDNCEYTVADVADKGYTLNIFAGKEKTPPEPKKEVTKANLIYADGKTQTAFFKGTFFKATSEAYAYADTLKKDNG 191
DB 215 YRVADALKKDNCEYTVADVADKGYTLNIFAGKEKTPPEPKKEVTKANLIYADGKTQTAFFKGTFFKATSEAYAYADTLKKDNG 274

QY 182 FKGTFFKATSEAYAYADTLKKDNGYTVADVADKGYTLNIFAGKEKTPPEPKKEVTKANLIYADGKTQTAFFKGTFFKATSEAYAYADTLKKDNG 241
DB 275 FKGTFFKATSEAYAYADTLKKDNGYTVADVADKGYTLNIFAGKEKTPPEPKKEVTKANLIYADGKTQTAFFKGTFFKATSEAYAYADTLKKDNG 334

QY 242 LIYADGKTQTAFFKGTFFKATSEAYAYADTLKKDNGYTVADVADKGYTLNIFAGKEKTPPEPKKEVTKANLIYADGKTQTAFFKGTFFKATSEAYAYADTLKKDNG 301
DB 335 LIYADGKTQTAFFKGTFFKATSEAYAYADTLKKDNGYTVADVADKGYTLNIFAGKEKTPPEPKKEVTKANLIYADGKTQTAFFKGTFFKATSEAYAYADTLKKDNG 394

QY 302 KPPEPMD-TYK--LILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVD-GEWYDDATK 357
DB 395 KPPEKQVTKENIYFDGTVQVATFGTFAEATAEAY--RYADLLSKEHGKVTADLEDG 452

QY 358 TFFVT-----EKPEVIDASELTPAVTTVKLVINGTKLGETTTKAVDAETAETAE 404
DB 453 GYINIRFAGKEPPEETPEPEVQD-----GYASVEAEAAKALKND 496

QY 405 KAFKQYANDNGVGVWY 422
DB 497 DVNKSVTIRQAGDGRYYY 514

RESULT 3

S54396
protein L precursor - Peptostreptococcus magnus (strain 3316)
C:Species: Peptostreptococcus magnus
A:Variety: strain 3316
C:Date: 27-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 15-Oct-1999
C:Accession: S54396
R:Murphy, J.P.; Duggleby, C.J.; Atkinson, M.A.; Trowern, A.R.; Atkinson, T.; Goward, C.H.

Mol. Microbiol. 12, 911-920, 1994
A:title: The functional units of a peptostreptococcal protein L.
A:Reference number: S54396; MUID:95020613
A:Accession: S54396
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-992 <MUR>
A:Cross-references: EMBL:L04466; MUID:g150673; PIDN:AAA67503.1; PID:g150674

Query Match 56.5%; Score 1263.5; DB 2; Length 992;
Best Local Similarity 58.8%; Pred. No. 4.9e-59;
Matches 281; Conservative 36; Mismatches 90; Indels 71; Gaps 12;

QY 2 VENKEETPETPDSEEEVTKANLIFANGSTQTAFFKGTFFKATSEAYAYADTLKKDNG 61
DB 241 MERKLEKETPE--PEEEVTKANLIFADGSTQNAEFKGFAYADAYADALAKDNG 298

QY 62 EYTVADVADKGYTLNIFAGKEKTPPEPKKEVTKANLIYADGKTQTAFFKGTFFKATSEAYAYADTLKKDNG 121
DB 299 EYTVADVADKGYTLNIFAGKEKTPPEPKKEVTKANLIYADGKTQTAFFKGTFFKATSEAYAYADTLKKDNG 358

QY 122 YRVADALKKDNCEYTVADVADKGYTLNIFAGKEKTPPEPKKEVTKANLIYADGKTQTAFFKGTFFKATSEAYAYADTLKKDNG 179
DB 359 YAYADLLAKENGEYTVADLEDGGNTINIKFAGKETPETPEPKKEVTKANLIFADGKTQTAFFKGTFFKATSEAYAYADTLKKDNG 418

QY 180 AEFKGTFFKATSEAYAYADLLAKENGEYTVADLEDGGNTINIKFAGKETPETPEPKKEVTKANLIFADGKTQTAFFKGTFFKATSEAYAYADTLKKDNG 237
DB 419 AEFKGTFFKATSEAYAYADLLAKENGEYTVADLEDGGNTINIKFAGKETPETPEPKKEVTKANLIFADGKTQTAFFKGTFFKATSEAYAYADTLKKDNG 478

QY 238 IKANLIYADGKTQTAFFKGTFFKATSEAYAYADLLAKENGEYTVADLEDGGNTINIKFAGKETPETPEPKKEVTKANLIFADGKTQTAFFKGTFFKATSEAYAYADTLKKDNG 297
DB 479 IKVNLIFADGKTQTAFFKGTFFKATSEAYAYADLLAKENGEYTVADLEDGGNTINIKFAGKETPETPEPKKEVTKANLIFADGKTQTAFFKGTFFKATSEAYAYADTLKKDNG 538

QY 298 K-----VDE-----KPPEP-----DNY-KLILNGKTLG----- 321
DB 539 EQDGNPGTITDEWLLKNKAEKELKZAGTSDLYFSLINKAKTVEGVKALNKLKA 598

QY 322 ---ET-----TTEAVDAATAEKVFKQYAND-----NGVDGEWYDDATKFTFTVK 364
DB 599 HAGEETPELKDGYATVEEAAKALKNDVNNAYEIVOGADGRYYY--VLKIEVADEE 656

QY 365 PEVIDASELTPAVTTVKLVINGTKLGETTTKAVDAETAETAEKAFKQYANDNGVGVWY 422
DB 657 EPGEDTPEVQEGYATYE-----EAAAKAEKALKEDKVNNAVYEVVQAGDGRYYY 704

RESULT 4

S00128
protein G precursor - Streptococcus sp. (Streptococcus G148)
N:Alternate names: albumin-binding protein; cell wall-bound protein
C:Species: Streptococcus sp.
A:Variety: Streptococcus G148
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 19-May-2000
C:Accession: S00128; A27604; A26314
R:Olsson, A.; Eliasson, M.; Guss, B.; Nilsson, B.; Hellman, U.; Lindberg, M.; Uhlen, Eur. J. Biochem. 168, 319-324, 1987
A:title: Structure and evolution of the repetitive gene encoding streptococcal protein G
A:Reference number: S00128; MUID:88029445
A:Accession: S00128
A:Molecule type: DNA
A:Residues: 1-593 <OLS>
A:Cross-references: EMBL:X06173; MUID:g47084; PIDN:CAA29540.1; PID:g47085
A:Note: the source is designated as Streptococcus G148
A:Note: part of this sequence, including the amino end of the mature protein, was cloned from a Streptococcus G148 strain
R:Sjoerbring, U.; Falkenberg, C.; Nielsen, E.; Akerstroem, B.; Bjoerck, L.
J. Immunol. 140, 1595-1599, 1988
A:title: Isolation and characterization of a 14-kDa albumin-binding fragment of streptococcal protein G
A:Reference number: A27604; MUID:88154455
A:Accession: A27604
A:Molecule type: protein
A:Residues: 62-101 <SJO>
R:Guss, B.; Eliasson, M.; Olsson, A.; Uhlen, M.; Frej, A.K.; Jornvall, H.; Flock, J.I. EMBO J. 5, 1567-1575, 1986

A;Title: Structure of the IgG-binding regions of streptococcal protein G.
A;Reference number: A26314; MUID:86300657
A;Accession: A26314
A;Molecule type: DNA
A;Residues: 114-593 <GUS>
A;Cross-references: GB:X04015; NID:g47071; PIDN:CAA27638.1; PID:g47072
C;Function:
C;Description: it is part of the cell wall structure of group G streptococci and is cov

C;Superfamily: M5 protein
C;Keywords: duplication; membrane protein
F;1-33/Domain: signal sequence #status predicted <SIG>
F;34-593/Product: protein G #status experimental <MAT>
F;34-116/Domain: alanine-rich <ALA>
F;117-290/Domain: AB duplication <Dup1>
F;117-140/Region: A repeat
F;141-191/Region: B repeat
F;192-215/Region: A repeat
F;216-266/Region: B repeat
F;267-290/Region: A repeat
F;303-497/Domain: IgG binding <IGB>
F;303-357/Region: C repeat
F;358-372/Region: D repeat
F;373-427/Region: C repeat
F;428-442/Region: D repeat
F;443-497/Region: C repeat
F;498-567/Domain: proline-rich <PRO>
F;568-593/Domain: carboxyl-terminal <CTD>

Query Match 31.7%; Score 708; DB 2; Length 593;
Best Local Similarity 48.3%; Pred. No. 3.7e-30;
Matches 205; Conservative 28; Mismatches 131; Indels 60; Gaps 17;

QY 45 ATSEAYAYADTLKNGEYTVADVADKGYTLNIFAGKEKTPPEKPEEVTIKANLIYA-DG 103
DB 101 ADALAKADALKEFN-KYGVSDYKRLNNAKTVGVKDLQAVVESAKKARISEATDG 159
QY 104 -----KTQTAEFKGTPEEATAEAYRYADALKKONGE---YTVADVADKGYTLNIFAGKEK 155
DB 160 LSDFLKSQTP-----AEDTVKSTELAEAKVLANRELDKYGVSDYHKNLNNAKTVEGVR 213
QY 156 TPEPKPEEVIKANLIYA-DG-----KTQTAEFKGTPEEATAEAYRYAD---LLAKENGK 206
DB 214 DLQAVVESAKKARISEATDGLSDFLKSQTP-----AEDTVKSTELAEAKVLANRELDK 267
QY 207 YTVADVADKGYTLNIFAGKEKTPPEE-----PKPEEVIKANLIYADKGTQTAEPKGTFAE 260
DB 268 YGVSDYKRLNNAKTVGVKDLQAVVESAKKARISEATDGLSDFLKSQTP-----LILNGKTLKE---FTTE 320
QY 261 A--TAEAYRYADLLAKENGKGTADLEDGGYTINIRFAGKVKDEKPE-----EPMDTY 310
DB 321 AVDAATAEKVKQYANDNG-----VDGEWYDDATKTFVTEKPEVIDASELTPAVTTY 374
QY 311 KLINGKTLGETTTEAVDAATAEKVKQYANDNGVGEWYDDATKTFVTEKPEVIDA 370
DB 375 KLINGKTLGETTTEAVDAATAEKVKQYANDNGVGEWYDDATKTFVTEKPEVIDA 434
QY 371 SELTPAVTTYKLINGKTLGETTTKAVDAETAETAEKAFKQYANDNGVGVWYDDATKTF 430
DB 435 SELTPAVTTYKLINGKTLGETTTKAVDAETAETAEKAFKQYANDNGVGVWYDDATKTF 494
QY 431 VTEM 434
DB 495 VTEM 498

RESULT 5
S42574
streptococcal surface protein - Streptococcus dysgalactiae
C;Species: Streptococcus dysgalactiae
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 19-May-2000
R;Jonsson, H.; Mueller, H.P.
Eur. J. Biochem. 220, 819-826, 1994

A;Title: The type-III Fc receptor from Streptococcus dysgalactiae is also an alpha(2)
A;Reference number: S42574; MUID:94192673
A;Accession: S42574
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-664 <ON>
A;Cross-references: EMBL:Z29666; NID:g470220; PIDN:CAA82764.1; PID:g470221
C;Superfamily: M5 protein

Query Match 31.6%; Score 706.5; DB 2; Length 664;
Best Local Similarity 43.9%; Pred. No. 5.2e-30;
Matches 192; Conservative 51; Mismatches 143; Indels 51; Gaps 17;

QY 24 ANLIFANGSTQTAETFEKATSEAYAYADTLK-KONGEYTVADVADKGYTLNIFAGKEK 79
DB 148 ANEIVNNSDAYTAESIOPLYKLINDAY--DVLESKDYSKY--DSQDKVNNLADQLRDV 202
QY 80 -----GKEKTPPEKPEEVIKANLIYADKGTQTAETFEKATSEAYAYADTLK-KONGEYTVADVADKGYTLNIFAGKEK 125
DB 203 QAVQLEAPTVIDAPELTPALTTVKLVKGNTF--SGETTTK-----AIDTATAEKEFKQYA 256
QY 126 DALKKONGEYTVADVADKGYTLNIFAGKEKTPPEKPEEVIKANLIYADKGTQTAETFE 182
DB 257 TANNVD-GEWYDDATKTFVTEKPAVIDAPELTPALTTVKLVKGNTF--SGETTTT--- 310
QY 183 KGTPEEATAEAYRYADLLAKENGKGYTVADVADKGYTLNIFAGKEKTPPEKPEEVIK 239
DB 311 KAVDAETAETAEKAFKQYATANNVDGEWYDDATKTFVTEKPAVIDAPELTPALTTVKLV 370
QY 240 ANLIYADKGTQTAETFEKATSEAYAYADLLAKENGKGTADLEDGGYTINIRFAGKVK 299
DB 371 GNIF--SGETTTKRAIDAATAEKEFKQYATANGV--DGEWYDDATKTFVTEKPA--V 422
QY 300 DEKPE--EPMDTYKLINGKTLGETTTEAVDAATAEKVKQYANDNGVGEWYDDATK 357
DB 423 IDAPELTPALTTVKLVKGNTFSGETTTKAVDAETAETAEKAFKQYANDNGVGEWYDDATK 482
QY 358 TFFVTEKPEVIDASELTPAVTTYKLINGKTLGETTTKAVDAETAETAEKAFKQYANDNGVD 417
DB 483 TFFVTEKPAVIDAPELTPALTTVKLVKGNTFSGETTTKAVDAETAETAEKAFKQYANDNGVD 542
QY 418 GVWYDDATKTFVTEM 434
DB 543 GVWYDDATKTFVTEM 559

RESULT 6
A24496
IgG-binding protein - Streptococcus sp. (group G)
C;Species: Streptococcus sp.
C;Date: 17-Sep-1987 #sequence_revision 17-Sep-1987 #text_change 17-Mar-2000
C;Accession: A24496; A39041
R;Fahnestock, S.R.; Alexander, P.; Nagle, J.; Filpula, D.
J. Bacteriol. 167, 870-880, 1986
A;Title: Gene for an immunoglobulin-binding protein from a group G Streptococcus.
A;Reference number: A24496; MUID:86304178
A;Accession: A24496
A;Molecule type: DNA
A;Residues: 1-448 <FAH>
A;Cross-references: GB:M13825; NID:g153822; PIDN:AAA03664.1; PID:g153823
R;Sjoeborg, U.; Bjoerck, L.; Kastern, W.
J. Biol. Chem. 266, 399-405, 1991
A;Title: Streptococcal protein G. Gene structure and protein binding properties.
A;Reference number: A39041; MUID:91093154
A;Accession: A39041
A;Status: preliminary
A;Molecule type: protein
A;Residues: 34-42, N', 45-48; 62-76; 186-200 <SJO>
C;Genetics:
A;Gene: spg
C;Superfamily: M5 protein
C;Keywords: transmembrane protein

Query Match 30.6%; Score 683; DB 2; Length 448;

Best Local Similarity 53.2%; Pred. No. 5.4e-29;

Matches 176; Conservative 15; Mismatches 92; Indels 48; Gaps 9;

QY 130 KNGEYTVDVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYAD-GKTQTAEFKGTTEE 188

Db 45 RKGELT-----NLGNSETTLALRNEESATADLTAAAVADTVAAAAENAGA 92

QY 189 ATAAYRYADLLA-----KENGKYTVDVADKGYTLNKFAGKEKTPPEPKKEVTIKAN 241

Db 93 AAWAAAAADALAKADALKEFNKYGVSYYKNLNNKTVGKIDLOAQVVESAKKAR 152

QY 242 LIYA-DG-----KTQTAEFKGTFAETAAYRYAD---LLAKENGKYTADLEDGGYINI 292

Db 153 ISEATDGLSDFLKSQTP-----AEDTVKSIELAAEKVLANRELDKYGV---SDYHKNL 202

QY 293 RFACKKYDEKPE-----EPMDTVKLLINGTKLKGTETTTAVDAATAEKVFKQYAND 343

Db 203 INNAKTVEGKELIDEILAAALPKTDYTKLILNGTKLKGTETTTAVDAATAEKVFKQYAND 262

QY 344 NGVDGEYTDYDANKTFVTVEKPEVIDASELTPAVTVYTKLVINGTKLKGTETTTAVDAETA 403

Db 263 NGVDGEYTDYDANKTFVTVEKPEVIDASELTPAVTVYTKLVINGTKLKGTETTTAVDAETA 322

QY 404 EKAFQYANDNGVDGYVWYTDATKTFVTTEM 434

Db 323 EKAFQYANDNGVDGYVWYTDATKTFVTTEM 353

RESULT 7

S55890

plasma protein receptor MAG precursor - Streptococcus dysgalactiae

C:Species: Streptococcus dysgalactiae

C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 17-Mar-2000

C:Accession: S55890

R:Jonsson, H.; Frykberg, L.; Rantamaeki, L.; Guss, B.

Gene 143, 85-89, 1994

A:Title: MAG, a novel plasma protein receptor from Streptococcus dysgalactiae.

A:Reference number: S55890; MUID:94259307

A:Accession: S55890

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-413 <JON>

A:Cross-references: EMBL:L27798; NID:g475117; PIDN:AAA26921.1; PID:g475118

C:Superfamily: M5 protein

Query Match 18.1%; Score 404; DB 2; Length 413;

Best Local Similarity 32.3%; Pred. No. 2.2e-14;

Matches 141; Conservative 31; Mismatches 110; Indels 154; Gaps 14;

QY 1 AVENKEETPETPETSDEEVTIKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKN 60

Db 30 AVNAEESTVSPVTATDAVT-----TSKEALAIINKLSEDN 66

QY 61 GEYTVDVADKGYTLNKFAGKEKTPPEPKKEVTIKANL-IYADGKTQTAEFKGTPEATA 119

Db 67 LN-NLDI-----QVLAKAGRIDLASDSADT--IKALLAEVTA 101

QY 120 EAYRYADALKDNGEYTVDVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQT 179

Db 102 EVTRL-----NEEKMARDAVDRAIAADAA-----AFSELKD 132

QY 180 AEFKGTPEETAAYRYADL-LAKENGKYTVDVADKGYTLNKFAGKEKTPPEPKKEVTI 238

Db 133 AOLKAYEDLAKLADTDLDLDAKIINDYTKVENAKTAEDVK-----KIFESQNEVT- 186

QY 239 KANLIYADGKTQTAEFKGTFAETAAYRYADLLAKENGKYTADLEDGGYINIRFAGKK 298

Db 187 -----RIKTERAKALAAALAKAKADAIE-----ILKKYG-----IGDYII----- 220

QY 299 VDEKPEEPMDTYKLLINGTKLKGTETTTAVDAATAEKVFKQYANDNGVDGEYTDATKT 358

Db 221 -----KLINGKTAEGVT-----ALKLD 237

QY 359 FVTEKPEVIDASELTPAVTVYTKLVINGTKLKGTETTTAVDAETAEKVFKQYANDNGVDG 418

Db 238 EILASKPAVIDAPELTPALTTVYTKLVINGTKLKGTETTTAVDAETAEKVFKQYANDNGVDG 297

QY 419 VWTYDDATKTFVTTEM 434

Db 298 VWTYDDATKTFVTTEM 313

RESULT 8

A42404

collagen adhesin - Staphylococcus aureus

C:Species: Staphylococcus aureus

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 10-Nov-1995

C:Accession: A42404; S27665

R:Pattil, J.M.; Jonsson, H.; Guss, B.; Switalski, L.M.; Wiberg, K.; Lindberg, M.; Hook

J. Biol. Chem. 267, 4766-4772, 1992

A:Title: Molecular characterization and expression of a gene encoding a Staphylococcus

A:Reference number: A42404; MUID:92165839

A:Contents: FDA 574

A:Accession: A42404

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1185 <PAT>

A:Cross-references: EMBL:M81736

A:Note: sequence extracted from NCBI backbone (NCBIP:83982)

Query Match 8.4%; Score 188.5; DB 2; Length 1185;

Best Local Similarity 23.8%; Pred. No. 0.016;

Matches 136; Conservative 61; Mismatches 166; Indels 209; Gaps 38;

QY 2 VENKEETPETPETSDEEVTIKANLIFANG-STQTAEFKGTPEKAT 46

Db 527 VYNTKEPIETTSISGEKVVWDDKQDGRKPEKVSVN---LLANGKVKVTLTLD-----VT 576

QY 47 SEA---YAYADTLKKNNG-----EYTVDVADKGYTLNKFAGKEKTPPEPKPE 90

Db 577 SETNMWYEFKDLPKYDEGKKIEYTVTEHDVHKDYTTDI--NGTTITNKY-----TPGESA 629

QY 91 EYV-----IKANLIYADGKTQ-----TAEFKGTPEEATAEAY 122

Db 630 TVTKNWDNNQDGRKPEIKVEL-YDQKATGKGTAILNESNNWTHWTGLDEKAKGQOV 688

QY 123 RY-ADALKKDNCEYT-VDVADKG-----YT-----LNKEFAGKEKTPPEPKKEVTIKA 168

Db 689 KYTVELTKVKGTYTHVDNNDGNLIVTNKYTPETTSISGEKVVWDDKQDGRKPEKVSV 748

QY 169 NLIYADG-KTQTAEFKGTPEEATAEA---YRYADLLAKENGK-----YTVDV 211

Db 749 NLL-ADGEKVKTL-----DVTSETNMWYEFKDLPKYDEGKKIEYTVTEHDVHKDYTTDI 800

QY 212 ADGKYTLNKFAGKEKTPPEPKKEVT-----IKANLIYADGKTQ----- 250

Db 801 --NGTTITNKY-----TPGETSATVTKNWDNNQDGRKPEIKVEL-YDQKATGKGTAI 852

QY 251 -----TAEFKGTFAETAEAAYRY-ADLLAKENGKYTADLED---GGYTIINIRF----- 294

Db 853 LNESNNWTHWTGLDEKAKGQOVKYTVTEELTKVKG-YTHVDNNDGNLIVTNKYTPETT 911

QY 295 --AGKKV-DEK-----PEEPMDTYKLLINGTKLKGTETTTAVDAATAEKVFKQYANDN 344

Db 912 SISGEKVVWDDKQDGRKPEKV--SVNLLANGKVKV---TLDVTSNWKYEFKDLPK-- 964

QY 345 GVDGEYTDATK-TFTVTEKPEVIDASELTPAVTVYTKLVINGTKLKGTETTTAVDAETA 403

Db 965 -----YDEGKKIEYTVTE-----DHVKDYTTDINGTTI---TNKYTPGETS 1002

QY 404 EKAFQYANDNGVDG-----VWTYDDATKT 428

Db 1003 AVTVKNWDNNQDGRKPEIKVELYDQKAT 1034

JC6171
late embryogenesis abundant protein - Arabidopsis thaliana
N;Alternate names: embryogenic cell protein 63; phosphotyrosine protein
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 21-Jul-2000
C;Accession: JC6171
R;Yang, H.; Saitou, T.; Kameda, Y.; Harada, H.; Kamada, H.
Gene 184, 83-88, 1997
A;Title: Arabidopsis thaliana ECP63 encoding a LEA protein is located in chromosome 4.
A;Reference number: JC6171; MUID: 97169149
A;Accession: JC6171
A;Molecule type: mRNA
A;Residues: 1-448 <YAN>
A;Cross-references: DDBJ:D64140; NID: g1526423; PIDN: BAA11017.1; PID: g1526424
C;Comment: This protein is a phosphotyrosine protein which belongs to late embryogenesis
C;Genetics:
A;Gene: ATECP63
A;Map position: 4, south part
C;Keywords: seed

Query Match 7.5%; Score 168; DB 2; Length 448;
Best Local Similarity 24.2%; Pred. No. 0.059;
Matches 113; Conservative 50; Mismatches 161; Indels 142; Gaps 20;
QY 3 ENKEETPETDSEEVTIKANLIFANGSTQTAFAKGTPEKATSEAYAYADTLKKNDE 62
DB 79 EAAESTKEGAIASEKAVGAK-----DATVEKAKET----ADYTAEKVGE 119
QY 63 YTVVDVADKGYTLNIFKAGKEKTPPEKPEEVIKANLIYADKGTQTAFFKGTFFPEATAEAY 122
DB 120 Y-----KDYTVDAKPAKOTTAETAKETANTYADKA-VEAKDKTAEKIGEYK----- 165
QY 123 RYADALKKNGEYTVVDVADKGYTLNIFKAGKEKTPPEKPEEVIKANLIYADKGTQTAFF 182
DB 166 -----DYADKA-----VEAKDKTAETAKETANTYADKA-KEAKDKTAEK 204
QY 183 KGTPEETAEE-----AYRYADLLAKNGKGYTVVDVADKGYTLNIFKAGKEKTP 230
DB 205 VGEYKDYTVDAKVEARDYTAETAKAIAEKDKTAETGEY-----KDYTVKATEGKDVTS 258
QY 231 ---EPKEEV--TKANLIYADKGTQTAFFKGTPEATAEAYRYADLLAKNGKGYTADLED 285
DB 259 KLGEKDSAVETAKRAMGFLSGTTEAKGKAVETKDT-----AKE----- 298
QY 286 GGYTINIRFAGKVDKPEPEPMDTYKLLNGKTLKGGTTTTEAVDAAT-----AEKV 336
DB 299 -----NMEKAGEVTRQKME-----MRLEGKELKEBAGAKAEASOKTRESTESGAOKA 347
QY 337 --FKYANDNGVDGEYTDYDATKTFVTTEKPEVIDASEL-----TPAVTTYKL 382
DB 348 EETKDSAPVRGNEAKGTFICGALGNVTETAIKSKLTMPDSIVETRAAREHGGTGVTVVEYK 407
QY 383 VINGKTLKGTTKYKAVDAETAEAKFQYANDNG--VDGVWYTDATK 427
DB 408 VEDSKPGKAVATSLKASQDMTG-----QTFNDVGRMD-----DDARK 443

RESULT 13
T52485
neurofilament protein NF-M(2), middle molecular weight [imported] - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000
C;Accession: T52485
R;Gervasi, C.; Szaro, B.G.
Brain Res. Mol. Brain Res. 48, 229-242, 1997
A;Title: Sequence and expression pattern of two forms of the middle molecular weight neu
A;Reference number: Z26090
A;Accession: T52485
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-913 <GER>
A;Cross-references: EMBL:U085970; PIDN: AAB53390.1

Query Match 7.2%; Score 162; DB 2; Length 913;
Best Local Similarity 24.4%; Pred. No. 0.29;
Matches 119; Conservative 57; Mismatches 172; Indels 140; Gaps 23;
QY 3 ENKEETPETDSEEVTIKANLIFANGSTQTA-----EFKGTPEKATSEAY 50
DB 474 EEEEGEETQGEEEAEIEVAAV---ESSVQAAPGEAEAEAEAEAEAEAEAE 530
QY 51 AVADTLKNGEYTVVDVADKGYTLNIFKAGKEKTPPEK-----EVTIKANLIYADGKT 105
DB 531 EKEEEKEEKE-----DEG-----EKVDEEEGGGQGEEDYKAE-----EKEE 571
QY 106 QTAFFKGTFFEEATAEAYRYADALKKNGEYTVVDVADKGYTLNIFKAGKEKTPPEKPEEY 165
DB 572 EGEEDAEVEETKABE-----GEEGEGEDEAEK-----EGSGEEKEEVA 614
QY 166 IYKANLIYADKGTQTAFFKGTFFEEATAEAYRYADLLAKE--NGKYTVVDVADKGYTLNIFKAG 224
DB 615 -----AD-----EGGEGEEGEGEGEKDEKDEKEGEAEAEAE-----GG 652
QY 225 KEKTPPEKPEEVIKANLIYADKGTQTAFFKGTFFA--EATAEAYRYADLLAKE--NGKY 279
DB 653 EBEETTEAAQAEVETMTETKIVREKAEVQTEAEDEETQDAEOKDEGEDEETGE 712
QY 280 TADLEDGGYTIINIRFAG--KKVDEKPEEPMDTYKLLNGKTLKG----- 321
DB 713 TEKEEGEKEEDADSGTEQEEKEAEKEGE--IDEGETKEGAEEDAGEDKQAKSP 769
QY 322 ETTTEAVDAATAEKVPKQYANDNGVDGEYTDYDATKTFV----TEKPEVIDASELT 377
DB 770 KSSEETISYKSAASVSGSEKAEKSTQKDDGKEDTAVNGDVEGEGVKESEDKAV 829
QY 378 TTYKLVING-----KTLKGETT---TKAVDAETAEKAFQYANDNGVDGVWYTD 425
DB 830 T-----NGLDSSPSDEKSSKGGQVTVTKKLETTTSE-----GEDGT---K 870
QY 426 TKTFVTTE 433
DB 871 TKSVTVTQ 878
RESULT 14
C95008
immunoglobulin A1 proteinase [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C;Accession: C95008
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapp
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A;Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID: 21357209; PMID: 11463916
A;Accession: C95008
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1856 <KUR>
A;Cross-references: GB:AE005672; PIDN: AAK74260.1; PID: g14971537; GSPDB: GN00164; TIGR:
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP0071
C;Superfamily: Streptococcus sanguis IgA-specific metalloendopeptidase
Query Match 7.1%; Score 159; DB 2; Length 1856;
Best Local Similarity 23.1%; Pred. No. 1;
Matches 119; Conservative 60; Mismatches 176; Indels 160; Gaps 28;
QY 7 ETPETDSEEVTIKANLIFANGSTQTAFFKGT-----FEKAT-SEAYAYAD---- 54
DB 315 ETKGTQEPCHGEAAVREDLPVY----TKPLETKGQPGHGEAAVREEPAYTEPLAT 370
QY 55 -----TLKKNNGEYTVVDVADKGYTLNIFKAGKEKTPPE--PKEEVTIK-----A 96

Db	371	KGTOEPCHGKATVREETLEYTEPVATKG--TQEPHEGEAAVEEELPALEVTTRNRTEIQ	429
Qy	97	NLIYADGKTO-----TASFKGTFEEATAEAYRADALKKDNGEYV-----VDVADG	142
Db	430	NIPYTEEIQDPTLLKNRRKIERQG---QACFTTQYEDYIVNGVNVETKEVSRTEVAPV	486
Qy	143	GYTINLKFAKKE-----TPEEKEEVTIKANLI-----YADGKTQTAEFKGTFFEA	189
Db	487	NEVVKGTLLVKVPTEITNLTKVENKKSITVSYNLIDTTSAYVSAKIQV--FHG-----	539
Qy	190	TAEAYRYADLLAKENGKYIVDVADKGYTLNINFAKGEKTPPEP--KEEV-----TI	238
Db	540	-----DKLVRE-----VDI-----ENPAKEQVIGSLDYTPYTV	568
Qy	239	KANLIYADG-----KTQTAEFKGTFAEATAEAYRYADLLAKENGKYTADLEDGGYTN	291
Db	569	KTHLTNLNGENNEENTETSTQDFQBYKKIEIKIDSVELYKENDRY-----	616
Qy	292	IRFAGKKVDEKPEEPMDYTKLLINGKT-----LKGETTTEAVDAATAEKFQYANDN	344
Db	617	-----RRYLSLSAPDTAKYFVKVKSDFRKEMYLPVKSITENTDGTGTVTVAVDQVLEE	671
Qy	345	GVDGWTYDDATKTFV-----TEKPEVIDASBELTPAVTYYKLIVNG--KTLKGETTTKAVD	399
Db	672	GTDG---YKD--DYTFVAKSKAEQPCGV--TSFQQLVYTAQMSNLSSVYTLASDMTADSV	725
Qy	400	AETAEAKFQYANDNGVDGWYTYDDATKTFVTTEM	434
Db	726	LGBQKQSYLTGAFTGSLIG-----SDGTSKYAYIDL	756

RESULT 15

B98047
hypothetical protein sprl403 [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: B98047
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; Dehoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: B98047
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2551 <RUR>
A:Cross-references: GB:AE007317; PIDN:AAL00207.1; PID:gn15459054; GSPDB:GN00174
C:Genetics:
A:Gene: sprl403

Query Match	7.1%:	Score 159;	DB 2;	Length 2551;
Best Local Similarity	23.0%:	Pred. No. 1.5;		
Matches 107;	Conservative	46;	Mismatches 162;	Indels 150;
Gaps 21;				

Qy	2	VENKETPETPDSSEVVIKANLIFANGSTQTAEFKGTFEKATSEAYADTLKKDNG	61
Db	1965	IDRGDSAPTITTKONGDGTHTITVQNDGSESTTVVKGDRDGT---	ANITTTENPDG 2020
Qy	62	EYTVDVADKGYTLNIFACKKTPPEEKV-----TJK-----	ANLIYADGKT 105
Db	2021	SHITVTNPDGSKETVWVNGDKGTPKVEYVDNDNDGTHTVKVTGCGNVTNALIKGDK	2080
Qy	106	QTAEFKGTFEETAAYRAYADALKKDNGEYTVDVADKGYTLNIFACKKTPPEEKBEVT	165
Db	2081	GKA-----ATATTENPD-----GSHTVTITN-----	PDGTKNEFV 2111
Qy	166	IKANLIYADGKQTAEFKGTFEETAAYRAYADLLAKENGKYTVDVADKGYTLNIFACKG	225
Db	2112	VKNRGDVGDRPTTAS-----VRDNG-----	DGSHITVI--- 2140

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:28:27 ; Search time 13.5074 Seconds
(without alignments)
784.805 Million cell updates/sec

Title: US-08-325-278B-3

Perfect score: 2235

Sequence: 1 AVENKEETPETDSEEV.....GVDGVWVYDDATKFTVTM 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2235	100.0	434	2	US-08-795-475-3
2	1565	70.0	305	2	US-08-795-475-1
3	1263.5	56.5	1027	4	US-08-446-137B-2
4	1216	54.4	291	4	US-08-446-137B-4
5	706.5	31.6	664	3	US-08-669-408B-2
6	622.5	27.9	502	1	US-08-378-761A-25
7	622.5	27.9	502	1	US-08-485-286-25
8	622	27.8	493	1	US-08-378-761A-23
9	622	27.8	493	1	US-08-485-286-23
10	618	27.7	342	3	US-08-828-741B-6
11	618	27.7	342	4	US-09-160-567-6
12	618	27.7	489	1	US-08-378-761A-19
13	618	27.7	489	1	US-08-485-286-19
14	615	27.5	178	3	US-08-828-741B-13
15	615	27.5	178	4	US-09-160-567-13
16	615	27.5	198	3	US-08-828-741B-8
17	615	27.5	198	4	US-09-160-567-8
18	614	27.5	495	3	US-08-828-741B-4
19	614	27.5	495	4	US-09-160-567-4
20	613	27.4	182	3	US-08-828-741B-2
21	613	27.4	182	4	US-09-160-567-2
22	404	18.1	413	3	US-08-669-408B-10
23	369.5	16.5	402	1	US-08-378-761A-17
24	369.5	16.5	402	1	US-08-485-286-17
25	334	14.9	75	4	US-08-446-137B-8
26	315.5	14.1	71	4	US-08-446-137B-6
27	306	13.7	74	4	US-08-446-137B-7

28	303	13.6	71	4	US-08-446-137B-5	Sequence 5, Appli
29	293	13.1	57	4	US-09-117-233-16	Sequence 16, Appl
30	292	13.1	57	4	US-09-117-233-2	Sequence 2, Appli
31	292	13.1	60	4	US-09-117-233-12	Sequence 12, Appl
32	287	12.8	56	4	US-09-058-459-38	Sequence 38, Appl
33	287	12.8	56	4	US-09-127-926-38	Sequence 38, Appl
34	286	12.8	57	4	US-09-117-233-8	Sequence 8, Appli
35	286	12.8	58	4	US-09-117-233-10	Sequence 10, Appl
36	286	12.8	58	4	US-09-444-305A-1	Sequence 1, Appli
37	286	12.8	60	4	US-09-117-233-6	Sequence 6, Appli
38	285	12.8	60	4	US-09-117-233-4	Sequence 4, Appli
39	283	12.7	56	4	US-09-058-459-68	Sequence 68, Appl
40	283	12.7	56	4	US-09-127-926-68	Sequence 68, Appl
41	281	12.6	56	4	US-09-058-459-46	Sequence 46, Appl
42	280	12.5	56	4	US-09-058-459-43	Sequence 43, Appl
43	280	12.5	56	4	US-09-058-459-44	Sequence 44, Appl
44	280	12.5	56	4	US-09-058-459-58	Sequence 58, Appl
45	280	12.5	56	4	US-09-058-459-70	Sequence 70, Appl

ALIGNMENTS

RESULT 1
US-08-795-475-3
; Sequence 3, Application US/08795475
; Patent No. 5965390
; GENERAL INFORMATION:
; APPLICANT: Bjvick, Lars
; APPLICANT: Sjvbring, Ulf
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,475
; FILING DATE: 11-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mcmasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.402DI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 434 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli LE392/pHDLG, DSM 7055
US-08-795-475-3

2
DMS
+
calc
#2
pdp

Query Match 100.0%; Score 2235; DB 2; Length 434;
Best Local Similarity 100.0%; Pred. No. 1.8e-170;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AVENKEETPETDSEEVTKANLIFANGSTQAEKFTFEKATSEAYADTLKKDN 60
|||||

Db 1 AVENKEETPETDSEEEVVIKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKN 60
Qy 61 GEYTVADVADKGYTLNIFAGKEKTPPEPKKEVVIKANLIYADGKTQTAEFKGTPEEATAE 120
Db 61 GEYTVADVADKGYTLNIFAGKEKTPPEPKKEVVIKANLIYADGKTQTAEFKGTPEEATAE 120
Qy 121 AYRYADALKDNGEYTVADVADKGYTLNIFAGKEKTPPEPKKEVVIKANLIYADGKTQTA 180
Db 121 AYRYADALKDNGEYTVADVADKGYTLNIFAGKEKTPPEPKKEVVIKANLIYADGKTQTA 180
Qy 181 EFKGTPEEATAEAYRYADLLAKENKGYTVADVADKGYTLNIFAGKEKTPPEPKKEVVIKA 240
Db 181 EFKGTPEEATAEAYRYADLLAKENKGYTVADVADKGYTLNIFAGKEKTPPEPKKEVVIKA 240
Qy 241 NLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENKGYTVADVADKGYTLNIFAGKEKTP 300
Db 241 NLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENKGYTVADVADKGYTLNIFAGKEKTP 300
Qy 301 EKPEPMDTYKLIILNGKTLKGETTTEAVDAATAEKFVKQYANDNGVDGWTYDDATKTFT 360
Db 301 EKPEPMDTYKLIILNGKTLKGETTTEAVDAATAEKFVKQYANDNGVDGWTYDDATKTFT 360
Qy 361 VTERPEVIDASELPVATYTKLVINGKTLKGETTTKAVDAETAETAKFKQYANDNGVDGVW 420
Db 361 VTERPEVIDASELPVATYTKLVINGKTLKGETTTKAVDAETAETAKFKQYANDNGVDGVW 420
Qy 421 TYDDATKTFTVTEM 434
Db 421 TYDDATKTFTVTEM 434

RESULT 2
US-08-795-475-1
; Sequence 1, Application US/08795475
; Patent No. 5965390
; GENERAL INFORMATION:
; APPLICANT: Bjvick, Lars
; APPLICANT: Sjvbring, Ulf
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,475
; FILING DATE: 11-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.402D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 305 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli LE392/pHDL, DSM 7054

US-08-795-475-1

Query Match 70.0%; Score 1565; DB 2; Length 305;
Best Local Similarity 100.0%; Pred. No. 2,6e-117;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AVENKEETPETDSEEEVVIKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKN 60
Db 1 AVENKEETPETDSEEEVVIKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKN 60
Qy 61 GEYTVADVADKGYTLNIFAGKEKTPPEPKKEVVIKANLIYADGKTQTAEFKGTPEEATAE 120
Db 61 GEYTVADVADKGYTLNIFAGKEKTPPEPKKEVVIKANLIYADGKTQTAEFKGTPEEATAE 120
Qy 121 AYRYADALKDNGEYTVADVADKGYTLNIFAGKEKTPPEPKKEVVIKANLIYADGKTQTA 180
Db 121 AYRYADALKDNGEYTVADVADKGYTLNIFAGKEKTPPEPKKEVVIKANLIYADGKTQTA 180
Qy 181 EFKGTPEEATAEAYRYADLLAKENKGYTVADVADKGYTLNIFAGKEKTPPEPKKEVVIKA 240
Db 181 EFKGTPEEATAEAYRYADLLAKENKGYTVADVADKGYTLNIFAGKEKTPPEPKKEVVIKA 240
Qy 241 NLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENKGYTVADVADKGYTLNIFAGKEKTP 300
Db 241 NLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENKGYTVADVADKGYTLNIFAGKEKTP 300
Qy 301 EKPEE 305
Db 301 EKPEE 305

RESULT 3
US-08-446-137B-2
; Sequence 2, Application US/08446137B
; Patent No. 6182903
; GENERAL INFORMATION:
; APPLICANT: Trowern, Angus R.
; APPLICANT: Atkinson, Anthony
; APPLICANT: Murphy, Jonathan P.
; APPLICANT: Laurence, Oliver S.
; APPLICANT: Duggleby, Clive J.
; TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED
; TITLE OF INVENTION: FROM L PROTEIN AND THEIR USES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,137B
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.406
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1027 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-446-137B-2

```

Query Match          56.5%; Score 1263.5; DB 4; Length 1027;
Best Local Similarity 58.8%; Pred. No. 1.5e-92;
Matches 281; Conservative 36; Mismatches 90; Indels 71; Gaps 12;

QY 2 VENKEETPETDSEEVITKANLIFANGSTQTAERKGTPEKATSEAYAYADTLKKDNG 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 276 MERKLEKETPE--PEEEVITKANLIFADGSTQNAEFKGTFAKAVSDAYADALKKDNG 333

QY 62 EYVDVADKGYTLNIFAGKEKTPPEEKVEVITKANLIYADGKTQTAEFGTPEEATAEA 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 334 EYVDVADKGLTLNIFAGKEKTPPEEKVEVITKANLIYADGKTQTAEFGTPEEATAEA 393

QY 122 YRADALKDNGEYTVVDVADKGYTLNIFAGKE--KTPPEEKVEVITKANLIYADGKTQT 179
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 394 YADLAKENGEYTDLEDGGNTINIKFAGKETPETPEEKVEVITKANLIYADGKTQT 453

QY 180 AEFKGTPEEATAYADLAKENGYTVVDVADKGYTLNIFAGKE--KTPPEEKVEVIT 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 454 AEFKGTPEEATAYADLAKENGEYTDLEDGGNTINIKFAGKETPETPEEKVEVIT 513

QY 238 IKANLIYADGKTQTAEFGTFAEATAYADLAKENGYTDLEDGGYTTINIRFAGK 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 514 IKVNLIFADGKTQTAEFGTFAEATAYADLAKENGYTDLEDGGYTTINIRFAGK 573

QY 298 K-----VDE-----KPEEPN-----DTY-KLILNGKTLKG----- 321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 574 EQGPNPGITIDEWLLKNAKEEAIKEUKAGITSLSLAKVTEGVGVEALKNEILKA 633

QY 322 -----ET-----TTEAVDAATAEKVKQYAND-----NGVDGEWYTDATKTFVTEK 364
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 634 HAGEETPELKGATYEEAEAAKALKNDDVNNAYEIVQAGDGRYYY--VLKIEVADEE 691

QY 365 PEVIDASELTPATVTKYLVLNGTKLGETTTKAVDAETAKEAFKQYANDNGVDGWYTY 422
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 692 EPGEDTPEVQEGYATY-----EAEAAKALKEKDKVNNAYEIVQAGDGRYYY 739

```

RESULT 4

```

US-08-446-137B-4
; Sequence 4, Application US/08446137B
; Patent No. 6162903
; GENERAL INFORMATION:
; APPLICANT: Trower, Angus R.
; APPLICANT: Atkinson, Anthony
; APPLICANT: Murphy, Jonathan P.
; APPLICANT: Laurence, Oliver S.
; APPLICANT: Dugleby, Clive J.
; TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED
; FROM L PROTEIN AND THEIR USES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,137B
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: McWaters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.406
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 291 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-446-137B-4

```

```

Query Match          54.4%; Score 1216; DB 4; Length 291;
Best Local Similarity 82.5%; Pred. No. 1.6e-89;
Matches 241; Conservative 20; Mismatches 25; Indels 6; Gaps 3;

QY 10 ETPEITDSEEVITKANLIFANGSTQTAERKGTPEKATSEAYAYADTLKKDNGEYTVVDVAD 69
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 ETPE--PEEEVITKANLIFADGSTQNAEFKGTFAKAVSDAYADALKKDNGEYTVVDVAD 59

QY 70 KGYTLNIFAGKEKTPPEEKVEVITKANLIYADGKTQTAEFGTPEEATAEAYRYADALK 129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 KGLTLNIFAGKEKTPPEEKVEVITKANLIYADGKTQTAEFGTPEEATAEAYRYADALLA 119

QY 130 KNGEYTVVDVADKGYTLNIFAGKE--KTPPEEKVEVITKANLIYADGKTQTAEFGTPE 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 KNGEYTDLEDGGNTINIKFAGKETPETPEEKVEVITKANLIYADGKTQTAEFGTPE 179

QY 188 EATAEAYRYADLAKENGYTVVDVADKGYTLNIFAGKE--KTPPEEKVEVITKANLIYA 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 180 EATAKAYAYANLAKENGEYTDLEDGGNTINIKFAGKETPETPEEKVEVITKANLIYA 239

QY 246 DGKTQTAEFGTFAEATAYADLAKENGYTDLEDGGYTTINIRFAGK 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 240 DGKTQTAEFGTFAEATAYADLAKENGYTDLEDGGYTTINIRFAGK 291

```

RESULT 5

```

US-08-669-408B-2
; Sequence 2, Application US/08669408B
; Patent No. 6100055
; GENERAL INFORMATION:
; APPLICANT: GUSS, Bengt
; APPLICANT: JONSSON, Hans
; APPLICANT: LINDBERG, Martin
; APPLICANT: MUELLER, Hans-Peter
; APPLICANT: RANTAMAKI, Liisa K.
; TITLE OF INVENTION: METHOD AND MEANS FOR PRODUCING
; PLASMA PROTEINASE INHIBITOR-BINDING PROTEINS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,408B
; FILING DATE: 03-JUL-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/SE94/00826
; FILING DATE: 06-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9302855-3
; FILING DATE: 06-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768

```

REFERENCE/DOCKET NUMBER: 61743/102
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 664 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-669-408B-2

Query Match 31.6%; Score 706.5; DB 3; Length 664;
Best Local Similarity 43.9%; Pred. No. 1.9e-48;
Matches 192; Conservative 51; Mismatches 143; Indels 51; Gaps 17;

QY 24 ANLIFANGSTQTAFFKGTFFKATSEAYAYADTLK-KDNGEYTVDVADKGYTL--NFKFA-79
Db 148 ANEIVNNSDAYTAESIPLYKLINDAY---DVLESKDYSKY--DSQDKVNNLADQLRDV 202
QY 80 -----GKEKTPPEPKKEVTIKANLIYADGKTOTAEFKGTFEATAEA--YRYA 125
Db 203 QAVOLEAPTVIDAPELTPALTTTKLVVKGNTF--SGETITK----AIDTATAKEFKQYA 256
QY 126 DALKKONGEYTVDVADKGYTLNFKFA---GKEKTPPEPKKEVTIKANLIYADGKTOTAEF 182
Db 257 TANNVD-GEWSYDDATKTFVTEKPAVIDAPELTPALTTTKLVKGNTP--SGETIT---310
QY 183 KGTFEATAYAYADLLAKENKGYTVDVADKGYTLNFKFA---GKEKTPPEPKKEVTIK 239
Db 311 KAVDAETAFAEKAFQYATANNVDGEWSYDDATKTFVTEKPAVIDAPELTPALTTTKLVK 370
QY 240 ANLIYADGKTOTAEFKGTFAEATAYAYADLLAKENKGYTADLEGGYTNIRFAGKKV 299
Db 371 GNTF--SGETITKAIDATAEKEFKQYATANGV---DGWSYDDATKTFVTEKPA---V 422
QY 300 DEKPE--EPMDYTKLILNGKLTGKETTTEAVDAATAEKVKFYANDNGVDGEWYDDATK 357
Db 423 IDAPELTPALTTTKLVKGNTEGTTTKAVDAETAFAEKAFQYANENGWYGEWSYDDATK 482
QY 358 TFFVTEKPEVIDASELTPAVTTTKLVINGKTLKGETTTKAVDAETAFAEKAFQYANDNGVD 417
Db 483 TFFVTEKPAVIDAPELTPALTTTKLVINGKTLKGETTTKAVDAETAFAEKAFQYANENGVD 542
QY 418 GVWYDDATKTFVTTEM 434
Db 543 GVWYDDATKTFVTTEM 559

RESULT 6
US-08-378-761A-25
Sequence 25, Application US/08378761A
Patent No. 5635384

GENERAL INFORMATION:
APPLICANT: WALSH, TERENCE A
APPLICANT: HEY, TIMOTHY D
APPLICANT: MORGAN, ALICE ER
TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE
TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A METHOD OF
TITLE OF INVENTION: USING
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSEE: ANDREA T. BORUCKI
STREET: 9330 ZIONSVILLE ROAD
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: US
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/378,761A
FILING DATE: 26-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BORUCKI, ANDREA T
REGISTRATION NUMBER: 33651
REFERENCE/DOCKET NUMBER: 38272B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 337-4846
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 502 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-378-761A-25

Query Match 27.9%; Score 622.5; DB 1; Length 502;
Best Local Similarity 41.1%; Pred. No. 6.3e-42;
Matches 178; Conservative 36; Mismatches 136; Indels 83; Gaps 16;

QY 37 EFKGTF-----EKATSEAYAYADTLKDNNGEYTVDV-ADKGYTLNFKAGKEKTPPEPK 89
Db 38 DHKGFQPVLPPEKKVPPELWFYTE-LKTRTSITLAIKRDNLVYLVGFRTPG-----87
QY 90 EEVTIKANLIYADGKTOTAEFKGTFEATAYAYADALKKDNNGEYTVDVADKGYTLNFK 149
Db 88 -----GVWVEFGDGDTHLLGDNPRWLGFGGRYOD-LIGNKGLTETVTMGRAETRAVN 139
QY 150 FAGKEKTPPEPKKEVTIKANLIYADGKTOTAEFKGTFEATAYAYADLLAKENKGYTV 209
Db 140 DLAKKKKAADPQ-----ADTKSLVK---LVWVCEGLRFNTV-----SRIV 178
QY 210 DV---ADKGYTLNFKAGKEKTPPEPKKEVTIKANLIYADGKTOT-AEPKGTGFABATA 265
Db 179 DAGFSQHSQVTLTFTV-OGK---QVQKWDRIKSAFAFWADHPTAVIPDMQKLGKIDKNEA 233
QY 266 YRYADLLAKENKGYTADLEGGYTNIRFAGKKVDEKPE-----EPMDYTKLILNGK 317
Db 234 ARIVALV--KNQTTAAATAGSSCARVRSSCGVD-KPEVIDASELTPAVTTTKLVINGK 290
QY 318 TLKGETTTTEAVDAATAEKVKFYANDNGVDGEWYDDATKTFVTTE-----363
Db 291 TLKGETTTTEAVDAATAEKVKFYANDNGVDGEWYDDATKTFVTTEKPEVIDASELTPAV 350
QY 364 ---KPEVIDASELTPAVTTTKLVINGKTLKGETTTKAVDAETAFAEKAFQYANDNGVDGW 420
Db 351 TRSKPEVIDASELTPAVTTTKLVINGKTLKGETTTTEAVDAATAEKVKFYANDNGVDGEW 410
QY 421 TYDDATKTFVTTE 433
Db 411 TYDDATKTFVTTE 423

RESULT 7
US-08-285-286-25
Sequence 25, Application US/08485286
Patent No. 5646026
Patent No. 5646026 5646119
GENERAL INFORMATION:
APPLICANT: WALSH, TERENCE A
APPLICANT: HEY, TIMOTHY D
APPLICANT: MORGAN, ALICE ER
TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE
TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD OF
TITLE OF INVENTION: USING
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSEE: ANDREA T. BORUCKI
STREET: 9330 ZIONSVILLE ROAD
CITY: INDIANAPOLIS


```
; TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE
; TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD OF
; TITLE OF INVENTION: USING
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ANDREA T. BORUCKI
; STREET: 9330 ZIONSVILLE ROAD
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,286
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/378761
; FILING DATE: 26-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BORUCKI, ANDREA T
; REGISTRATION NUMBER: 33651
; REFERENCE/DOCKET NUMBER: 38272B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 337-4846
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 493 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-485-286-23

Query Match 27.8%; Score 622; DB 1; Length 493;
Best Local Similarity 46.9%; Pred. No. 6.7e-42;
Matches 146; Conservative 28; Mismatches 55; Indels 82; Gaps 9;

QY 195 RYADLLAKENKGYTVADVADKGYTLNFKAGKEKTPPEKPEV---TIKANLIYADG---- 247
Db 114 RYQDLNKG--GLETVTMGRAEMTRAVNDLAKKKKADPDQADTKSKLVKLVVMVCEGLRFN 172
QY 248 ---KTQTAEFKG-----TFEAATAEAYRYAD---: : : : : 270
Db 173 TVSRITVDAGFNSQHGVTLTITQCKQVQKWDRIKAAFEWADHPTAVIPDMQKLGIDKNE 232
QY 271 ---LLAKENGYTADLEDGGYTTINIRPAGKKVDEKPE-----EPMDTYKLILANGKTL 319
Db 233 AARIVALVKQNTAAATAAG-SYNNV-----DKPEVIDASELTPAVTTYKLVLINGKTL 283
QY 320 KGETTTEAVDAATAEKVKFYQYANDNGVDGWTYDDATKFTTVE----- 363
Db 284 KGETTTEAVDAATAEKVKFYQYANDNGVDGWTYDDATKFTTVEKPEVIDASELTPAVTR 343
QY 364 -KPEVIDASELTPAVTTYKLVLINGKTLKGETTKAVDAETAERAKFYQYANDNGVDGWTY 422
Db 344 SKPEVIDASELTPAVTTYKLVLINGKTLKGETTEAVDAATAEKVKFYQYANDNGVDGWTY 403
QY 423 DDAKTKFTTVE 433
Db 404 DDAKTKFTTVE 414

RESULT 10
US-08-828-741B-6
; Sequence 6, Application US/08828741B
; Patent No. 6043069
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Suess, Gabriele M.
```

```
; APPLICANT: Tarlinton, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; TITLE OF INVENTION: PRODUCING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,741B
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-828-741B-6

Query Match 27.7%; Score 618; DB 3; Length 342;
Best Local Similarity 75.9%; Pred. No. 8.4e-42;
Matches 126; Conservative 5; Mismatches 15; Indels 20; Gaps 1;

QY 16 SEEVTIKANLIPANGSTQTAEFKGTFFKATSEAYADTLKKDNGEYTVADVADKGYTLN 75
Db 174 SAFEVTIKANLIPANGSTQTAEFKGTFFKATSEAYADTLKKDNGEYTVADVADKGYTLN 233
QY 76 IKFAGKEKTPPE-----PKEVTIKANLIYADGKTQTAEFKGTPE 115
Db 234 IKFAGKEATNRNTDGSYDGYILOINRWGLTSAAEVTIKANLIFANGSTQTAEFKGTPE 293
QY 116 EATAEAYRYADALKKNGEYTVADVADKGYTLNFKAGKEKTPPEPK 161
Db 294 KATSEAYADTLKKDNGEYTVADVADKGYTLNFKAGKESAWRHPQ 339

RESULT 11
US-09-160-567-6
; Sequence 6, Application US/09160567
; Patent No. 6326179
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Suess, Gabriele M.
; APPLICANT: Tarlinton, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; TITLE OF INVENTION: PRODUCING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/160,567
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/828,741
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-160-567-6

```

```

Query Match 27.7%; Score 618; DB 4; Length 342;
Best Local Similarity 75.9%; Pred. No. 8.4e-42;
Matches 126; Conservative 5; Mismatches 15; Indels 20; Gaps 1;

QY 16 SEEVTIKANLIFANGSTQTAEPKGFPEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 75
Db 174 SAEVTIKANLIFANGSTQTAEPKGFPEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 233

QY 76 IKFAGKEKTPPEE-----PKEEVTIKANLIYADGKTQTAEFKGTPE 115
Db 234 IKFAGKEKTPPEE-----PKEEVTIKANLIYADGKTQTAEFKGTPE 293

QY 116 EATAEAYRAYADALKKNGEYTVDVADKGYTLNKKFAGKEKTPPEPK 161
Db 294 KATSEAYAYADTLKKDNGEYTVDVADKGYTLNKKFAGKESAWRHPQ 339

```

```

RESULT 12
US-08-378-761A-19
; Sequence 19, Application US/08378761A
; Patent No. 5635384
; GENERAL INFORMATION:
; APPLICANT: WALSH, TERENCE A
; APPLICANT: HEY, TIMOTHY D
; APPLICANT: MORGAN, ALICE ER
; TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE
; TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD OF
; TITLE OF INVENTION: USING
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ANDREA T. BORUCKI
; STREET: 9330 ZIONSVILLE ROAD
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/378,761A
; FILING DATE: 26-Jan-1995
; CLASSIFICATION: 435

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: BORUCKI, ANDREA T
; REGISTRATION NUMBER: 33651
; REFERENCE/DOCKET NUMBER: 38272B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 337-4846
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 489 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-378-761A-19

Query Match 27.7%; Score 618; DB 1; Length 489;
Best Local Similarity 40.8%; Pred. No. 1.4e-41;
Matches 175; Conservative 35; Mismatches 131; Indels 88; Gaps 16;

QY 37 EFKGTF-----EKATSEAYAYADTLKKDNGEYTVDV-ADKGYTLNKKFAGKEKTPPEPK 89
Db 38 DHKGIFQPVLPPEKKVPELWFYTE-LKTRTSSITLAIRMDNLVGVGFTPG-----87

QY 90 EEVTIKANLIYADGKTQTAEFKGTPEEATAEAYRAYADALKKNGEYTVDVADKGYTLN 149
Db 88 -----GVWWEFGKDGDTLLGDGNPRWLGFGGRYQD-LIGNKGLFTVTMGRAMTRAVN 139

QY 150 FAGKEKTPPEKPEVTIKANLIYADGKTQTAEFKGTPEEATAEAYRAYADLLAKENGKTYV 209
Db 140 DLAKKKKADPQ-----ADTKSKLVK-----LVWVCEGLRENTV-----SRTV 178

QY 210 DV---ADKGYTLNKKFAGKEKTPPEKPEVTIKANLIYADGKTQT-AEFGKTFEAEATAEA 265
Db 179 DAGFNSOHGVTLVLT-QGK---QVQKWDRIKAAFEWADHPAVIPDMQKLGKDKNEA 233

QY 266 YRVADLLAKENGKYTADLEDGGYTIINIRFAGKK---VDEKPEEP-MDTYKLIINGKTLKG 321
Db 234 ARTVALVKNQ-----TAAATAGSKPEVIDASELTPAVTYYKLVINGKTLKG 281

QY 322 ETTTEAVDAATAEKVKQYANDNGVDGWTYDDATKTFTVTE-----K 364
Db 282 ETTTEAVDAATAEKVKQYANDNGVDGWTYDDATKTFTVTEKPEVIDASELTPAVTRSK 341

QY 365 PEVIDASELTPAVTYYKLVINGKTLKGTTTKAVIDAETAEKAFKQYANDNGVDGWTYDD 424
Db 342 PEVIDASELTPAVTYYKLVINGKTLGTTTTEAVDAATAEKVKQYANDNGVDGWTYDD 401

QY 425 ATKTFVTPE 433
Db 402 ATKTFVTPE 410

RESULT 13
US-08-485-286-19
; Sequence 19, Application US/08485286
; Patent No. 5646026
; Patent No. 5646026 5646119
; GENERAL INFORMATION:
; APPLICANT: WALSH, TERENCE A
; APPLICANT: HEY, TIMOTHY D
; APPLICANT: MORGAN, ALICE ER
; TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE
; TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD OF
; TITLE OF INVENTION: USING
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ANDREA T. BORUCKI
; STREET: 9330 ZIONSVILLE ROAD
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,286
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/378761
; FILING DATE: 26-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BORUCKI, ANDREA T
; REGISTRATION NUMBER: 33651
; REFERENCE/DOCKET NUMBER: 38272B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 337-4846
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 489 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-485-286-19

```

```

Query Match 27.7%; Score 618; DB 1; Length 489;
Best Local Similarity 40.8%; Pred. No. 1.4e-41;
Matches 175; Conservative 35; Mismatches 131; Indels 88; Gaps 16;

```

```

QY 37 EFKGTF-----EKATSEAYAYADTLKNGEYTDV-ADKGYTLNKFAGKEKTPPEPK 89
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38 DHKGIFQVLPPEKKVPELWFYTE-LKTRTSSITLAIRMDNLVGFRTFG----- 87
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 90 EEVTIKANLIYADGKTQTAEFKGTFFEEATAEAYADALKKNGEYTDVADKGYTLNFK 149
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 88 -----GVWWEFGKGDTHLLGDNFWLFGGGRYQD-LIGNKGLFVTVMGREMTFVAVN 139
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 150 FAGEKTPPEPKPEVITKANLIYADGKTQTAEPKGTFFEEATAEAYADLLAKENGKTYV 209
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 140 DLAKKKAAADPQ-----ADTKSLVK---LVWVCEGLRENTV-----SRTV 178
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 210 DV---ADKGYTLNKFAGKEKTPPEKPEVITKANLIYADGKTQT-AEFKGTFAETAFA 265
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 DAGFNSQHGVTLAVT-OGK---OVQKWDRIKAAFEWADHPHTAVIPDMQKLGKDKNEA 233
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 266 YRVADLLAKENGKTYADLDGGYTNIRFAGKK---VDEKPEEP-MDTYKLIINGKTLKG 321
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 ARIVALVKNT-----TAAATAGSKPEVIDASELTPAVTTTKLVINGKTLKG 281
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 322 ETTEAVDAATAEKVKQYANDNGVDGEWYDDATKFTVTE-----K 364
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 282 ETTTEAVDAATAEKVKQYANDNGVDGEWYDDATKFTVTEKPEVIDASELTPAVTRSK 341
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 365 PEVIDASELTPAVTTTKLVINGKTLGETTTKAVDAETAFAKQYANDNGVDGVTYDD 424
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 342 PEVIDASELTPAVTTTKLVINGKTLGETTTKAVDAETAFAKQYANDNGVDGVTYDD 401
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 425 ATKFTVTE 433
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 402 ATKFTVTE 410
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 14
US-08-828-741B-13
; Sequence 13, Application US/08828741B
; Patent No. 6043069
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Suess, Gabriele M.
; APPLICANT: Tarlington, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; NUMBER OF SEQUENCES: 14

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,741B
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: DIGIGLIO, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 178 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-828-741B-13

```

```

Query Match 27.5%; Score 615; DB 3; Length 178;
Best Local Similarity 76.2%; Pred. No. 5.9e-42;
Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1;

```

```

QY 18 EEVTIKANLIYADGKTQTAEFKGTFFEEATAEAYADTLKNGEYTDVADKGYTLNFK 77
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12 EEVTIKANLIYADGKTQTAEFKGTFFEEATAEAYADTLKNGEYTDVADKGYTLNFK 71
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 78 FAGEKTPPE-----PKEVITKANLIYADGKTQTAEFKGTFFEEA 117
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72 FAGEKATNTDGTDDYGLQINRWGGLTSABEVTIKANLIYADGKTQTAEFKGTFFEEA 131
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 118 TAEAYADALKKNGEYTDVADKGYTLNFKFAGKEKTPPEPK 161
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132 TSEAYAYADTLKNGEYTDVADKGYTLNFKFAGKESAWRHPQ 175
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 15
US-09-160-567-13
; Sequence 13, Application US/09160567
; Patent No. 6326179
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Suess, Gabriele M.
; APPLICANT: Tarlington, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/160,567
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/828,741
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 178 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-160-567-13

```

```

Query Match      27.5%; Score 615; DB 4; Length 178;
Best Local Similarity 76.2%; Pred. No. 5.9e-42;
Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1;

QY 18 EEVTIKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKNNGEYTVDVADKGYTLNIK 77
Db 12 EEVTIKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKNNGEYTVDVADKGYTLNIK 71
QY 78 FAGKEKTPPEE-----PKEVTIKANLIYADGKTQTAEFKGTPEEA 117
Db 72 FAGKEATNRNTDGDYGIILQINSRWGGLTSAEVTIKANLIFANGSTQTAEFKGTPEKA 131
QY 118 TAEAYRYADALKKNNGEYTVDVADKGYTLNIKFKAGKEKTPPEPK 161
Db 132 TSEAYAYADTLKKNNGEYTVDVADKGYTLNIKFKAGKESAWRHPQ 175

```

Search completed: October 29, 2002, 09:31:57
Job time : 16.5074 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:25:06 : Search time 34.0622 Seconds
(without alignments)
1415.233 Million cell updates/sec

Title: US-08-325-278B-3

Perfect score: 2235

Sequence: 1 AVENKEETPETDSEEV.....GVDGVWYDDATKTFVTTEM 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11107396 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802:*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2235	100.0	434	14	Sequence encoding
2	1565	70.0	305	14	Immunoglobulin lig
3	1550	69.4	467	22	Amino acid sequenc
4	1372	61.4	367	21	Expression vector
5	1263.5	56.5	1027	14	Protein L. Peptoc
6	1263.5	56.5	1027	14	Protein L. Peptoc
7	1216	54.4	291	14	Immunoglobulin bin
8	706.5	31.6	664	16	S. dysgalactiae MI
9	702	31.4	593	15	Streptococcus Prot
10	701.5	31.4	594	12	Streptococcus GX78
11	692.5	31.0	480	8	Sequence of polype

12	692	31.0	593	11	AA07014	Protein G variant
13	683	30.6	448	8	AA070493	Protein G. Strept
14	683	30.6	448	10	AA095030	Protein G. Strept
15	683	30.6	448	11	AA07013	Protein G variant
16	683	30.6	448	12	AA01004	Streptococcus GX78
17	683	30.6	448	15	AA053290	Streptococcus GX78
18	675	30.2	269	15	AA053295	IgG-binding Strept
19	674	30.2	265	12	AA010011	Type 4 GX7809 prot
20	674	30.2	265	15	AA053294	IgG-binding Strept
21	670	30.0	185	21	AA057610	Streptococcus stra
22	665	29.8	235	10	AA094785	Protein G variant.
23	665	29.8	235	11	AA07004	Protein G variant.
24	663	29.7	144	10	AA094784	Protein G variant.
25	663	29.7	144	11	AA07003	Protein G variant.
26	663	29.7	144	12	AA010009	Type 3 GX7809 prot
27	663	29.7	144	15	AA053293	IgG-binding Strept
28	663	29.7	164	10	AA094787	Protein G variant.
29	663	29.7	164	11	AA07011	Protein G variant.
30	663	29.7	164	12	AA010015	Type 9 GX7809 prot
31	663	29.7	164	15	AA053299	IgG-binding Strept
32	663	29.7	199	10	AA094783	Protein G variant.
33	663	29.7	199	11	AA07002	Protein G variant.
34	663	29.7	199	12	AA010007	Type 2 GX7809 prot
35	663	29.7	199	15	AA053292	IgG-binding Strept
36	663	29.7	201	12	AA010013	Type 8 GX7809 prot
37	663	29.7	201	15	AA053298	IgG-binding Strept
38	658	29.4	148	10	AA094782	Protein G variant.
39	658	29.4	148	11	AA07001	Protein G variant.
40	658	29.4	148	12	AA010008	Type 1 GX7809 prot
41	658	29.4	148	15	AA053291	IgG-binding Strept
42	657	29.4	201	11	AA07010	Protein G variant.
43	656	29.4	269	10	AA094788	Protein G variant.
44	648	29.0	269	11	AA07006	Protein G variant.
45	646	28.9	201	10	AA094790	Protein G variant.

ALIGNMENTS

RESULT 1
AA042994

ID AAR42994 standard; Protein: 434 AA.

XX AC AAR42994:

XX DT 16-MAY-1994 (first entry)

XX Sequence encoding immunoglobulin light chain binding protein.

XX DE Immunoglobulin; light chain; binding; identification; purification;
XX KW separation; ss.

XX OS E. coli L392/pHDLG, DSM 7055.

XX Key Location/Qualifiers

XX FH Domain 5..305

XX FT /label= B1 immunoglobulin light chain binding domain.

XX FT Domain 81..305

XX FT /label= B2 immunoglobulin light chain binding domain.

XX FT Domain 153..305

XX FT /label= B3 immunoglobulin light chain binding domain.

XX FT Domain 225..305

XX FT /label= B4 immunoglobulin light chain binding domain.

XX FT Domain 297..305

XX FT /label= B5 immunoglobulin light chain binding domain.

XX FT Domain 309..434

XX FT /label= C1 immunoglobulin heavy chain binding domain.

FT Domain 364..434
FT /label= D intermediate immunoglobulin heavy
FT chain binding domain.
FT Domain 379..434
FT /label= C2 immunoglobulin heavy chain binding
FT domain.
XX WO9322342-A.
XX
XX
PD 11-NOV-1993.
XX
XX 28-APR-1993; 93WO-SE00375.
XX
XX 28-APR-1992; 92SE-0001331.
XX
XX (HIGH-) HIGHTECH RECEPTOR AB.
XX
XX Bjoerck L, Sjoerbring U;
XX
XX WPI; 1993-368722/46.
XX P-PSDB; AAR42994.
XX
XX New protein L binding light chains of all immunoglobulin classes
PT - for binding purifying and identifying immunoglobulin, also
PT related DNA, vectors and host cells
XX
XX Claim 6; Page 39-40; 71pp; English.
XX
XX Protein L (AAR42993) is capable of binding to immunoglobulin G light
CC chains. It is useful for binding, separating (purifying) and
CC identifying immunoglobulin and for removing immunoglobulin molecules
CC from serum. This is the coding sequence of one hybrid protein of
CC the L protein. The hybrid proteins can bind all human
CC immunoglobulin classes and many immunoglobulins from other species.
CC They are highly soluble and retain their binding activity at high
CC temperatures over a pH range of 3-10. They can be immobilised
CC without loss of activity.
XX
XX Sequence 434 AA;
XX
Query Match 100.0%; Score 2235; DB 14; Length 434;
Best Local Similarity 100.0%; Pred. No. 1.1e-123;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AVENKEETPETDSEEVITKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKN 60
DB 1 AVENKEETPETDSEEVITKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKN 60
QY 61 GEYTVADVADKGYTLNKFAGKEKTPPEPKBEVITKANLIYADGKTQTAPEKGFEEATAE 120
DB 61 GEYTVADVADKGYTLNKFAGKEKTPPEPKBEVITKANLIYADGKTQTAPEKGFEEATAE 120
QY 121 AYRYADALKDNGEYTVADVADKGYTLNKFAGKEKTPPEPKBEVITKANLIYADGKTQTA 180
DB 121 AYRYADALKDNGEYTVADVADKGYTLNKFAGKEKTPPEPKBEVITKANLIYADGKTQTA 180
QY 181 EFKGTPEATAEAYRYADLLAKENGKGYTVADVADKGYTLNKFAGKEKTPPEPKBEVITKA 240
DB 181 EFKGTPEATAEAYRYADLLAKENGKGYTVADVADKGYTLNKFAGKEKTPPEPKBEVITKA 240
QY 241 NLIYADGKTQTAPEKGFEEATAEAYRYADLLAKENGKGYTVADLEDGGYTNIRFAGKKVD 300
DB 241 NLIYADGKTQTAPEKGFEEATAEAYRYADLLAKENGKGYTVADLEDGGYTNIRFAGKKVD 300
QY 301 EKPEPMDTVKLLINGKTLKGTTTEAVDAATAEAYRYADLLAKENGKGYTVADLEDGGYTNIR 360
DB 301 EKPEPMDTVKLLINGKTLKGTTTEAVDAATAEAYRYADLLAKENGKGYTVADLEDGGYTNIR 360
QY 361 VTEKPEVIDASELTPAVTTYKLVINGKTLKGTTTAVDAATAEAYRYADLLAKENGKGYTVAD 420
DB 361 VTEKPEVIDASELTPAVTTYKLVINGKTLKGTTTAVDAATAEAYRYADLLAKENGKGYTVAD 420
QY 421 TYDDATKTFVTTEM 434

DB 421 TYDDATKTFVTTEM 434
RESULT 2
AAR42993
ID AAR42993 standard; Protein; 305 AA.
XX
XX AAR42993;
XX
XX 16-MAY-1994 (first entry)
XX
XX Immunoglobulin light chain binding protein (protein L).
XX Immunoglobulin; light chain; binding; identification; purification;
XX separation.
XX
XX E. coli LE392/pHDL, DSM 7054.
XX
XX Key Location/Qualifiers
FH 5..305
FT Domain /label= B1 immunoglobulin light chain binding
FT domain.
FT Domain 81..305
FT /label= B2 immunoglobulin light chain binding
FT domain.
FT Domain 153..305
FT /label= B3 immunoglobulin light chain binding
FT domain.
FT Domain 225..305
FT /label= B4 immunoglobulin light chain binding
FT domain.
FT Domain 297..305
FT /label= B5 immunoglobulin light chain binding
FT domain.
XX WO9322342-A.
XX
XX 11-NOV-1993.
XX
XX 28-APR-1993; 93WO-SE00375.
XX
XX 28-APR-1992; 92SE-0001331.
XX
XX (HIGH-) HIGHTECH RECEPTOR AB.
XX
XX Bjoerck L, Sjoerbring U;
XX
XX WPI; 1993-368722/46.
XX N-PSDB; AAQ50452.
XX
XX New protein L binding light chains of all immunoglobulin classes
PT - for binding purifying and identifying immunoglobulin, also
PT related DNA, vectors and host cells
XX
XX Claim 1; Page 36; 71pp; English.
XX
XX The protein (protein L) is capable of binding to immunoglobulin G
CC light chains. It is useful for binding, separating (purifying) and
CC identifying immunoglobulin and for removing immunoglobulin molecules
CC from serum. Hybrid proteins of the L protein can bind all human
CC immunoglobulin classes and many immunoglobulins from other species.
CC They are highly soluble and retain their binding activity at high
CC temperatures over a pH range of 3-10. They can be immobilised
CC without loss of activity.
XX
XX Sequence 305 AA;
XX
Query Match 70.0%; Score 1565; DB 14; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.4e-84;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AVENKEETPETDSEEVITKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKN 60

|||||
Db 1 AVENKEETPTETDSEEEVTIKANLIFANGSTQTAEFKGTFEKATSEAYADTLKKN 60
QY 61 GEYTVADVADKGYTLNIFKAGKEKTPPEPKEEVTIKANLIYADGKTQTAEPKGTPEEATAE 120
Db 61 GEYTVADVADKGYTLNIFKAGKEKTPPEPKEEVTIKANLIYADGKTQTAEPKGTPEEATAE 120
QY 121 AYRYADALKKDNGEYTVADVADKGYTLNIFKAGKEKTPPEPKEEVTIKANLIYADGKTQTA 180
Db 121 AYRYADALKKDNGEYTVADVADKGYTLNIFKAGKEKTPPEPKEEVTIKANLIYADGKTQTA 180
QY 181 EFKGTPEEATAEAYRYADLLAKENGYTVADVADKGYTLNIFKAGKEKTPPEPKEEVTIKA 240
Db 181 EFKGTPEEATAEAYRYADLLAKENGYTVADVADKGYTLNIFKAGKEKTPPEPKEEVTIKA 240
QY 241 NLIYADGKTQTAEPKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTTINIRFAGKKVD 300
Db 241 NLIYADGKTQTAEPKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTTINIRFAGKKVD 300
QY 301 EKPEE 305
Db 301 EKPEE 305
RESULT 3
AAB31372
ID AAB31372 standard; Protein; 467 AA.
AC AAB31372;
XX
XX
DT 20-APR-2001 (first entry)
XX
XX Amino acid sequence of protein L/CBD cex/ER retaining peptide fusion.
DE
XX Protein production; food processing; protein antibiotic; feed enzyme;
KW Protein L: CBD cex protein; cell signal peptide.
XX
XX Synthetic.
OS
PN WO20007174-A1.
XX
XX 21-DEC-2000.
XX
XX 07-JUN-2000; 2000WO-IL00330.
XX
XX 10-JUN-1999; 99US-0329234.
XX
XX (CBDT-) CBD TECHNOLOGIES LTD.
PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
XX Shani Z, Shoseyov O;
PI
XX
XX WPI; 2001-112219/12.
DR N-PSDB; AAF24730.
XX
XX Expressing and isolating recombinant protein in a plant, useful for
PT producing large quantities of recombinant proteins, by expressing a
PT fusion protein including a cellulose binding peptide fused to a
PT recombinant protein
XX
XX Example; Fig 2a; 87pp; English.
PS
XX The specification describes a method for expressing and isolating
CC a recombinant protein in a plant. The method comprising expressing a
CC fusion protein including the recombinant protein and a cellulose
CC binding peptide fused to it, where the fusion protein is
CC compartmentalised and sequestered within plant cells, plant
CC tissue or cultured plant cells. The method is useful for obtaining large
CC quantities of the recombinant proteins and protein products in a simple
CC and cost-effective manner. Recombinant proteins may be used commercially,
CC such as in the food processing industry, e.g. glucoamylases and glucose
CC isomerases are used for converting starch to high fructose corn syrup,
CC proteinases for the hydrolysis of high molecular weight proteins and in

CC manufacturing leather or alcoholic beverages, pectinesterases for
CC pectin hydrolysis in food industry, lipases for cleaving ester linkage
CC in triglycerides, and for effluent treatment. The recombinant proteins
CC may further be used to produce protein antibiotics, which can be used
CC in healing processes, and to produce animal feed enzymes. The present
CC sequence represents a fusion protein of the invention, and comprises a
CC fusion of a cell signal peptide, protein L, CBD cex, and an endoplasmic
CC reticulum retaining peptide.
XX
SQ Sequence 467 AA;
Query Match 69.4%; Score 1550; DB 22; Length 467;
Best Local Similarity 79.2%; Pred. No. 1.8e-83;
Matches 328; Conservative 3; Mismatches 43; Indels 40; Gaps 7;
QY 1 AVENKEETPTETDSEEEVTIKANLIFANGSTQTAEFKGTFEKATSEAYADTLKKN 60
Db 40 AVENKEETPTETDSEEEVTIKANLIFANGSTQTAEFKGTFEKATSEAYADTLKKN 99
QY 61 GEYTVADVADKGYTLNIFKAGKEKTPPEPKEEVTIKANLIYADGKTQTAEPKGTPEEATAE 120
Db 100 GEYTVADVADKGYTLNIFKAGKEKTPPEPKEEVTIKANLIYADGKTQTAEPKGTPEEATAE 159
QY 121 AYRYADALKKDNGEYTVADVADKGYTLNIFKAGKEKTPPEPKEEVTIKANLIYADGKTQTA 180
Db 160 AYRYADALKKDNGEYTVADVADKGYTLNIFKAGKEKTPPEPKEEVTIKANLIYADGKTQTA 219
QY 181 EFKGTPEEATAEAYRYADLLAKENGYTVADVADKGYTLNIFKAGKEKTPPEPKEEVTIK 239
Db 220 EFKGTPEEATAEAYRYADLLAKENGYTVADVADKGYTLNIFKAGKEKTPPEPKEEVTIK 279
QY 240 ANLIYADGKTQTAEPKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTTINIRFAGKKV 299
Db 280 ANLIYADGKTQTAEPKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTTINIRFAGKKV 339
QY 300 DEKPEEPMQYTKLILNGKTLKGTTEAVDAA-----TAEKVFQYQYANDN 344
Db 340 DEKPEGIPPT-----PTPTSASGPAGCQVLWGVNQMNTGTAQVTVKN-TCGA 386
QY 345 GVDGENTYDDATKFTVTKEPEVIDASELT-----PAVTTYKLVIINGKTLKGET 393
Db 387 PVDG-WTL-----TFSPSQOQVQWSTQSGSAVTVRNAPWNGNIPAGGT 434
RESULT 4
AAB10432
ID AAB10432 standard; Protein; 367 AA.
XX
XX AAB10432;
XX
DT 01-DEC-2000 (first entry)
XX
XX Expression vector pSEX11L4 protein G.
DE
XX
XX Expression vector; antibody binding protein; monoclonal antibody; Neo-R;
KW B lymphocyte; myeloma cell; hybridoma; protein G; bla protein.
XX
XX Synthetic.
XX
PN DE19900635-A1.
XX
XX 13-JUL-2000.
XX
XX 11-JAN-1999; 99DE-1000635.
XX
XX 11-JAN-1999; 99DE-1000635.
XX
XX (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
XX Breitling F, Poustka A, Moldenhauer G;
PI WPI; 2000-499832/45.
XX
XX N-PSDB; AAA71428.
DR

XX Selecting monoclonal antibodies, by expressing them on the surface of
PT hybridomas attached to antibody-binding protein, then reaction with
PT antibody library -
XX
XX Claim 16; Fig 1; 22pp; German.
XX
XX This invention describes a novel method for the selection of monoclonal
CC antibodies (MAB) which comprises (i) fusing B lymphocytes with myeloma
CC cells to produce antibody-producing hybridomas such that the antibodies
CC are presented at the surface of the hybridomas by an antibody-binding
CC protein (I); and (ii) binding the antibody to antigens (Ag). The
CC invention also describes antibody-binding proteins (I) that comprise a
CC combination of the signal peptide of a murine immunoglobulin (Ig) kappa
CC chain or a murine MHC (major histocompatibility complex) Class I k(k)
CC molecule; an antibody-binding site of proteins A, G, L or LG, and the
CC transmembrane domain of PDGFR (platelet-derived growth factor receptor)
CC or CD52. The method is used to select MAB with specificity for particular
CC antigens. MAB can be selected without separate culture of hybridomas,
CC and selection can be made against many antigens in a library, optionally
CC on the basis of strength of affinity for a particular antigen. Complex
CC mixtures of hybridomas can be used for selection, reducing the time and
CC cost involved in MAB selection. This sequence represents the protein G,
CC contained in the expression vector pSEX114 which contains the protein G,
CC Neo-R and the bla protein described in the method of the invention.
XX
XX Sequence 367 AA;

Query Match 61.4%; Score 1372; DB 21; Length 367;
Best Local Similarity 82.4%; Pred. No. 3.6e-73;
Matches 271; Conservative 19; Mismatches 29; Indels 10; Gaps 3;
QY 5 KEETPEPTDSSEEVITKANLIFANGSTOTAEFGTFEKATSEAYAYADTLKKDNGEYT 64
DB 25 KEKTPPEP-----KEEVITKANLIYADGKTQTAEFGTFEATAEAYADALAKKDNGEYT 80
QY 65 VDADVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAEFGTFEATAEAYRY 124
DB 81 VDADVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAEFGTFEATAEAYRY 140
QY 125 ADALAKKDNGEYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAEFGK 184
DB 141 ADALAKKDNGEYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAEFGK 200
QY 185 TFEATAEAYRYADLAKKNGEYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIY 244
DB 201 TFEATAEAYRYADLAKKNGEYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIY 260
QY 245 ADGKTQTAEFGKTFEATAEAYRYADLAKKNGEYTVDVADKGYTLNIFAGKKTVDKPE 304
DB 261 ADGKTQTAEFGKTFEATAEAYRYADLAKKNGEYTVDVADKGYTLNIFAGKKTVDKPE 320
QY 305 EPMDTYKLIINGTKLGETTTEAVDAATA 333
DB 321 SEED-----LNG-AVDQNDTSQTSPSA 343

RESULT 5
AAR42203
ID AAR42203 standard; Protein; 1027 AA.
XX
AC AAR42203;
XX
DT 18-MAY-1994 (first entry)
XX
DE Protein L.
XX
XX Peptide; immunoglobulin; binding; analysis; purification; ELISA;
KW enzyme linked immunoabsorbant assay.
XX
XX Peptococcus magnus.
OS
XX
XX Location/Qualifiers
FH

Peptide 36..59
/label= Signal sequence.
Protein 60..968
/label= Mature protein L.
XX
XX W09322439-A.
XX
XX 11-NOV-1993.
XX
XX 07-MAY-1993; 93WO-CB00950.
XX
XX 07-MAY-1992; 92GB-0009804.
XX
XX 24-DEC-1992; 92GB-00026928.
XX
XX (PUBL-) PUBLIC HEALTH LAB SERVICE BOARD.
XX
XX Atkinson A, Duggleby CJ, Murphy JP, Trowern AR;
XX
XX WPI; 1993-368798/46.
XX
XX N-PSDB; AAQ50946.
XX
XX New immunoglobulin binding proteins derived from Protein L -
XX which bind immunoglobulin kappa light chains but not albumin or
XX cell walls
XX
XX Disclosure; Figure 1; 28pp; English.
XX
XX The synthetic immunoglobulin binding proteins derived from protein
XX L correspond to the repeated sequences in protein L which bind
XX immunoglobulin kappa light chains. They can be used in protein
XX analysis, purification procedures and other biochemical processes e.
XX g. ELISA. The synthetic molecules are of particular advantage if
XX they are free of regions in protein L which exhibit albumin and cell
XX wall binding.
XX
XX Sequence 1027 AA;
QY Query Match 56.5%; Score 1263.5; DB 14; Length 1027;
Best Local Similarity 58.8%; Pred. No. 2.7e-66;
Matches 281; Conservative 36; Mismatches 90; Indels 71; Gaps 12;
QY 2 VENKETPEPTDSSEEVITKANLIFANGSTOTAEFGTFEKATSEAYAYADTLKKDNG 61
DB 276 MEKLSKEKTPPE-----PEEVITKANLIYADGKTQTAEFGTFEATAEAYADALAKKDNG 333
QY 62 EYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAEFGTFEATAE 121
DB 334 EYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAEFGTFEATAE 393
QY 122 YRYADALAKKDNGEYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQT 179
DB 394 YAYADLAKKNGEYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQT 453
QY 180 AEFKGTFEATAEAYRYADLAKKNGEYTVDVADKGYTLNIFAGKEKTPPEPKEEVT 237
DB 454 AEFKGTFEATAEAYRYADLAKKNGEYTVDVADKGYTLNIFAGKEKTPPEPKEEVT 513
QY 238 IKANLIYADGKTQTAEFGTFEATAEAYRYADLAKKNGEYTVDVADKGYTLNIFAGK 297
DB 514 IKVNLIFADGKTQTAEFGTFEATAEAYRYADLAKKNGEYTVDVADKGYTLNIFAGK 573
QY 298 K-----VDE-----KPEEPM-----DTY-KLILNGKTLK----- 321
DB 574 EQDPENPGITIDENLKNKEAEIKELKEAGITSDLYFSLINAKKATVEGVEALKNEILKA 633
QY 322 -----ET-----TTEAVDAATAEFVKQYAND-----NGVDGEYTVDDATFTFTVEK 364
DB 634 HAGEETPELKDGATYATYEEAAAEAAKALKNDDVNNAYEIVQAGDGRYY--VLKIEVADEE 691
QY 365 PEVIDASELTPATVTVTKLVINGCKTLKGETTTKAVDAETAEKAFKQYANDNGVDGVV 422
DB 692 EPGEDTPEVOEGYATY-----EAAAEAAKALKEKDVNNAYEVQAGDGRYY 739

[illegible]

QY 130 KNGEYTVADVADKGYTLNIFKAGKE--KTPPEPKKEVTIKANLIYADGKTQTAEFKGTPE 187
Db 120 KNGEYTDADGGNTINIKFAGKETPEPEPKKEVTIKVNLIFADGKIQTAEFKGTPE 179
QY 188 EATAEAYRYADLLAKENGYTVADVADKGYTLNIFKAGKE--KTPPEPKKEVTIKANLIYA 245
Db 180 EATAKAYAYANLLAKENGYTADLEGGNTINIKFAGKETPEPEPKKEVTIKVNLIFA 239
QY 246 DGKTQTAEFKGTAEATAEAYRYADLLAKENGYTADLEGGYTTINIRFAGK 297
Db 240 DGKTQTAEFKGTAEATAEAYRYADLLAKVNGEYTDADLEGGYTTINIKFAGK 291

RESULT 8
AAR71929
ID AAR71929 standard; Protein; 664 AA.
XX
AC AAR71929;
XX
DT 22-SEP-1995 (first entry)
XX
DE S. dysgalactiae MIG.
XX
KW MIG; fast alpha-2-macroglobulin binding protein; FAM;
KW Plasma proteinase-inhibitor binding protein.
XX
OS Streptococcus dysgalactiae strain SC1.
XX
FH Key Location/Qualifiers
FT Peptide 1..30
FT Domain 209..278 /label= Sig_peptide
FT Domain 279..348 /label= IgG1
FT Domain /note= "IgG binding domain 1"
FT Domain /label= IgG2
FT Domain 349..418 /note= "IgG binding domain 2"
FT Domain /label= IgG3
FT Domain 419..488 /note= "IgG binding domain 3"
FT Domain /label= IgG4
FT Domain 489..558 /note= "IgG binding domain 4"
FT Domain /label= IgG5
FT Region 559..638 /note= "IgG binding domain 5"
FT Region 630..635 /note= "cell wall spanning region"
FT Peptide 639..664 /note= "putative wall anchoring motif"
FT Region /note= "membrane spanning domain"

XX
PN W09507296-A.
XX
XX 16-MAR-1995.
XX
XX 06-SEP-1994; 94WO-SE00826.
XX
XX 06-SEP-1993; 93SE-0002855.
XX
XX (GUSS/) GUSS B.
XX (JONS/) JONSSON H.
XX (LIND/) LINDBERG M.
XX (MUEL/) MUELLER H.
XX (RANT/) RANTAMAKI L K.
XX
XX Guss B, Jonsson H, Lindberg M, Mueller H, Rantamaki LK;
XX WPI; 1995-123382/16.
XX N-PSDB; AAQ89196.
XX
XX DNA encoding fast alpha 2-macro:globulin-binding proteins - used

PT to obtain prods. for sepn., detection or quantification or for
XX binding inhibition
PS Disclosure; Page 31; 50pp; English.
XX
CC A phage lambda GEM-11 library of S. dysgalactiae SC1 DNA was analyzed
CC for fast-acting alpha-2-macroglobulin (FAM)- and IgG-binding
CC activity. One clone, which expressed both activities, was
CC analyzed to obtain DNA encoding the FAM-binding protein, MIG
CC (AAR71929). The mig gene is given in AAQ89196.
XX
SQ Sequence 664 AA;
Query Match 31.6%; Score 706.5; DB 16; Length 664;
Best Local Similarity 43.9%; Pred. No. 7 4e-34;
Matches 192; Conservative 51; Mismatches 143; Indels 51; Gaps 17;
QY 24 ANLIFANGSTQTAEFKGTSEAYAYADTLK-KDNGEYTVADVADKGYTL--NIKFA- 79
Db 148 ANEIVNNSDAYTAESIPLYKLINDAY---DVLESKDYSKY--DSQDKVNNLADQLRDV 202
QY 80 -----GKEKTPPEPKKEVTIKANLIYADGKTQTAEFKGTFEATAEA--YRYA 125
Db 203 QAVQLEAPTVIDAPELTPTALTYKLVKGNTF--SGETTK-----AIDTATAEKEFKQYA 256
QY 126 DALKKDNGEYTVADVADKGYTLNIKFA---GKEKTPPEPKKEVTIKANLIYADGKTQTAEF 182
Db 257 TANNVD-GEWSYDDATKTFVTVEKPAVIDAPELTPTALTYKLVKGNTF--SGETTT--- 310
QY 183 KGTFEATAEAYRYADLLAKENGYTVADVADKGYTLNIFKA---GKEKTPPEPKKEVTIK 239
Db 311 KAVDAETAFAKQYATANNVDGEWSYDDATKTFVTVEKPAVIDAPELTPTALTYKLVK 370
QY 240 ANLIYADGKTQTAEFKGTAEATAEAYRYADLLAKENGYTADLEGGYTTINIRFAGK 299
Db 371 GNTF--SGETTTKAIDAATAEFAKQYATANGV---DGWSYDDATKTFVTVEKPA---V 422
QY 300 DEKPE--EPMDTYKLILNGKLTGETTTEAVIDAATAEKFVKQYANGVDGWTYDDATK 357
Db 423 IDAPELTPTALTYKLVKGNTFSGETTTKAVDAETAFAKQYANGVYGEWSYDDATK 482
QY 358 TFTVTEKPEVIDASELTPTAVTYKLVINGKTLKGETTTTAVDAETAFAKQYANGVD 417
Db 483 TFTVTEKPAVIDAPELTPTALTYKLVINGKTLKGETTTTAVDAETAFAKQYANGVD 542
QY 418 GWTYDDATKTFVTTEM 434
Db 543 GWTYDDATKTFVTTEM 559
RESULT 9
AAR62944
ID AAR62944 standard; Protein; 593 AA.
XX
AC AAR62944;
XX
DT 10-JAN-1995 (first entry)
XX
DE Streptococcus Protein G derived from strain GX7805.
XX
KW Streptococcus Protein G; variant; IgG binding activity;
KW Immunoglobulin; Lancefield Group G; bacterial Fc receptor.
XX
OS Streptococcus sp. GX7805.
XX
FH Key Location/Qualifiers
FT Domain 106..140 /label= A1
FT Region 141..178 /label= a1
FT FT /note= "linking region"
FT 179..215 /label= A2
FT

QY 293 RFAGKKVDEKPE-----EPMDTYKLLINGKTLKGETTTTAVDAATAEKVFKQYAND 343
 Db 203 INNAKTVEGVKELIDELAALPKTDYTKLLINGKTLKGETTTTAVDAATAEKVFKQYAND 262
 QY 344 NGVDGWTYDDATKFTVTTEKPEVIDASELTPAVTTYKLVINGKTLKGETTTTAKVADAETA 403
 Db 263 NGVDGWTYDDATKFTVTTEKPEVIDASELTPAVTTYKLVINGKTLKGETTTTAKVADAETA 322
 QY 404 EKAFKQYANDNGVDGWTYDDATKFTVTTEM 434
 Db 323 EKAFKQYANDNGVDGWTYDDATKFTVTTEM 353

RESULT 14
 AAP95030
 ID AAP95030 standard; protein; 448 AA.
 XX
 AC AAP95030;
 XX
 DT 04-JUL-1990 (first entry)
 XX
 DE Protein G.
 XX
 KW Protein G; immunoglobulin; Fc receptor; ds.
 XX
 OS Streptococcus sp.
 XX
 PN WO8810306-A.
 XX
 PD 29-DEC-1988.
 XX
 PF 20-JUN-1988; 88WO-US02084.
 XX
 PR 19-JUN-1987; 87US-0063959.
 XX
 PA (GENE-) GENEX CORP.
 XX
 PI Fahnstock SR;
 XX
 DR WPI; 1989-023848/03.
 DR N-PSDB; AAN91093.
 XX

Cloned protein G variant genes -
 PT expressing proteins having immunoglobulin-binding properties of
 PT protein G and derived from Streptococcus sp.
 XX
 PS Disclosure; ; 116pp; English.
 XX
 CC Protein G of non-pathogenic streptococcus and variants may be isolated,
 CC useful as bacterial Fc receptors eg in purification and detection of Abs.
 CC screening of hybridoma clones and treatment of disease.
 XX
 SQ Sequence 448 AA;

Query Match 30.6%; Score 683; DB 10; Length 448;
 Best Local Similarity 53.2%; Pred. No. 1.1e-32;
 Matches 176; Conservative 15; Mismatches 92; Indels 48; Gaps 9;

QY 130 KNGEYTVADVADKGYTLNKFAGKEKTPPEPKEEVIKANLIYAD-GKTQTAEFKGTTEE 188
 Db 45 RNGGELT-----NLLGNSETTLALRNEESATADLTAAAVADTVAAAAENAGA 92

QY 189 ATAFAEYADLLA-----KENGKYTVADVADKGYTLNKFAGKEKTPPEPKEEVIKAN 241
 Db 93 AAMEAAAAALAKAKADALKEFNKGVSDYYKLNINNAKTVEGIDKDLQAVVYESAKKAR 152

QY 242 LIVA-DG-----KTQTAEFKGTFAEATAEAYRYAD---LLAKENGKYTADLEDGGYTTNI 292
 Db 153 ISATDGLSDFLKSQTP-----AEDTVKSIELAEAKVLANRELDKYGV-----SDYHNNL 202

QY 293 RFAGKKVDEKPE-----EPMDTYKLLINGKTLKGETTTTAVDAATAEKVFKQYAND 343
 Db 203 INNAKTVEGVKELIDELAALPKTDYTKLLINGKTLKGETTTTAVDAATAEKVFKQYAND 262

Db 203 INNAKTVEGVKELIDELAALPKTDYTKLLINGKTLKGETTTTAVDAATAEKVFKQYAND 262
 QY 344 NGVDGWTYDDATKFTVTTEKPEVIDASELTPAVTTYKLVINGKTLKGETTTTAKVADAETA 403
 Db 263 NGVDGWTYDDATKFTVTTEKPEVIDASELTPAVTTYKLVINGKTLKGETTTTAKVADAETA 322
 QY 404 EKAFKQYANDNGVDGWTYDDATKFTVTTEM 434
 Db 323 EKAFKQYANDNGVDGWTYDDATKFTVTTEM 353

RESULT 15
 AAR07013
 ID AAR07013 standard; protein; 448 AA.
 XX
 AC AAR07013;
 XX
 DT 17-JAN-1991 (first entry)
 XX
 DE Protein G variant with two active sites.
 XX
 KW Immunoglobulin.
 XX
 OS Streptococcus sp. Lancefield Group G strain.
 XX
 FH Key Location/Qualifiers
 FT Active-site 228..282
 FT Active-site /label=B1
 FT Active-site 298..352
 FT Active-site /label=B2
 XX
 PN US4956296-A.
 XX
 PD 11-SEP-1990.
 XX
 PF 20-JUN-1988; 88US-0209236.
 XX
 PR 20-JUN-1988; 88US-0209236.
 PR 14-FEB-1986; 86US-0829354.
 PR 23-APR-1986; 86US-0854887.
 PR 17-FEB-1987; 87WO-US00329.
 PR 19-JUN-1987; 87US-0063959.
 XX
 PA (GENE-) GENEX CORP.
 XX
 PI Fahnstock SR;
 XX
 DR WPI; 1990-297491/39.
 DR N-PSDB; AAQ06018.
 XX
 PT Recombinant Protein G variants - obtd. using a cloned gene
 PT encoding Protein G from Streptococcus sp., used for binding
 PT immunoglobulin.
 XX
 PS Disclosure; Fig 8a-c; 48pp; English.
 XX
 CC Fragments and variants of the sequence are claimed esp. where
 CC incorporated into a non-pathogenic host eg. E.coli, and expressed
 CC at high levels.
 CC The variants have a higher binding efficiency and capacity for
 CC immunoglobulin, and may be used for purifying, detecting and
 CC isolating antibodies.
 XX
 SQ Sequence 448 AA;

Query Match 30.6%; Score 683; DB 11; Length 448;
 Best Local Similarity 53.2%; Pred. No. 1.1e-32;
 Matches 176; Conservative 15; Mismatches 92; Indels 48; Gaps 9;

QY 130 KNGEYTVADVADKGYTLNKFAGKEKTPPEPKEEVIKANLIYAD-GKTQTAEFKGTTEE 188
 Db 45 RNGGELT-----NLLGNSETTLALRNEESATADLTAAAVADTVAAAAENAGA 92

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3

Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:25:41 ; Search time 29,364 Seconds
(without alignments)
2556.865 Million cell updates/sec

Title: US-08-325-278b-3

Perfect score: 2235

Sequence: 1 AVENKEETPETDSEEEV.....GVDGWTYDDATKTFVTVM 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17294929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2235	100.0	455	2 Q53291	Q53291 streptococ
2	1587.5	71.0	719	2 Q51912	Q51912 peptostrept
3	1263.5	56.5	992	2 Q51918	Q51918 peptostrept
4	706.5	31.6	664	2 Q53975	Q53975 streptococ
5	697.5	31.2	669	2 Q93FM8	Q93FM8 streptococ
6	670	30.0	185	2 Q54181	Q54181 streptococ
7	481	21.5	429	2 Q56212	Q56212 streptococ
8	404	18.1	443	2 Q53974	Q53974 streptococ
9	294	13.2	60	2 Q53337	Q53337 streptococ
10	182.5	8.2	3198	5 Q9UBG8	Q9UBG8 manduca sex
11	182	8.1	479	10 Q9LF88	Q9LF88 arabidopsis
12	177.5	7.9	2402	2 Q9AER7	Q9AER7 staphylococ
13	175	7.8	448	10 Q9SKP0	Q9SKP0 arabidopsis
14	170.5	7.6	1849	2 Q9SAK2	Q9SAK2 lactobacill
15	170	7.6	2276	2 Q93TY6	Q93TY6 staphylococ
16	168	7.5	448	10 Q96246	Q96246 arabidopsis

```
17 162.5 7.3 2353 2 P71401
18 162 7.2 913 13 Q13099
19 161 7.2 2692 4 Q9Y503
20 160.5 7.2 5458 5 Q90459
21 160 7.2 2691 4 Q9NS12
22 160 7.2 2705 4 Q95303
23 160 7.2 2725 4 Q9NYE5
24 159 7.1 1856 16 Q97T80
25 158 7.1 2045 16 Q9A0K5
26 157.5 7.0 1166 2 Q86489
27 157 7.0 798 13 Q90307
28 156.5 7.0 564 2 Q69188
29 156.5 7.0 581 2 Q33742
30 156.5 7.0 663 2 Q30874
31 156.5 7.0 701 2 Q9KK48
32 156.5 7.0 969 5 Q9ND19
33 156 7.0 1368 2 Q93M90
34 154.5 6.9 1110 13 Q91255
35 154.5 6.9 1822 2 Q07290
36 154 6.9 737 16 Q9JYK4
37 154 6.9 845 5 Q9Y1P8
38 154 6.9 1698 2 Q9LC00
39 153.5 6.9 492 2 Q56193
40 153.5 6.9 573 2 Q56192
41 153.5 6.9 701 2 Q9RQT5
42 153.5 6.9 1873 2 Q9Z4N7
43 153 6.8 897 13 Q13098
44 152 6.8 3147 5 Q17464
45 151.5 6.8 717 2 Q9EY85
```

ALIGNMENTS

RESULT 1

```
Q53291
ID Q53291 PRELIMINARY; PRT; 455 AA.
AC Q53291:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PROTEIN LG (FRAGMENT).
OS Streptococcus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93094283; PubMed=1460053;
RA Kihlberg B.M., Sjoberg U., Kastern W., Bjorck L.;
RT Protein LG: a hybrid molecule with unique immunoglobulin binding
RT properties.;
RL J. Biol. Chem. 267:25583-25588(1992).
DR EMBL; S50809; AAA03280.1;
DR HSP; P06654; IPGX.
DR InterPro; IPR003147; B1.
DR InterPro; IPR000724; IGG_bind_B.
DR Pfam; PF02246; B1; 4.
DR Pfam; PF01378; IGG_binding_B; 2.
FT NON_TER 455 455
SQ SEQUENCE 455 AA; 49926 MW; 381FC235BBC8307B CRC64;
```

Query Match 100.0%; Score 2235; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 1e-100;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AVENKEETPETDSEEEVTKANLIFANGSTQTAFFKGFEEKATSEAYAYADTLKKDN 60
DB 22 AVENKEETPETDSEEEVTKANLIFANGSTQTAFFKGFEEKATSEAYAYADTLKKDN 81
QY 61 GEYTVDAVKGYTLNKKFAGKEKTPPEPKFEVTKANLIYADGKTQTAFFKGFEEATAE 120
DB 82 GEYTVDAVKGYTLNKKFAGKEKTPPEPKFEVTKANLIYADGKTQTAFFKGFEEATAE 141
```

THIS PAGE BLANK (USPTO)

OY 121 AYRYADALKKONGEYTVDVADKGYTLNIRFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 180
|||||
Db 142 AYRYADALKKONGEYTVDVADKGYTLNIRFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 201
|||||
OY 181 EFKGTEETAEAYRYADLLAKENGKGYTVDVADKGYTLNIRFAGKEKTPPEPKEEVTIKA 240
|||||
Db 202 EFKGTEETAEAYRYADLLAKENGKGYTVDVADKGYTLNIRFAGKEKTPPEPKEEVTIKA 261
|||||
OY 241 NLITYADGKTQTAEFKGTFAEAYRYADLLAKENGKGYTVADLEGGYTNIRFAGKKYD 300
|||||
Db 262 NLITYADGKTQTAEFKGTFAEAYRYADLLAKENGKGYTVADLEGGYTNIRFAGKKYD 321
|||||
OY 301 EKPEEPMGYTKLILNGKTKLGETTTEAVDAATAEKYVFKOYANDNGVDGEMTYDDATKFTT 360
|||||
Db 322 EKPEEPMGYTKLILNGKTKLGETTTEAVDAATAEKYVFKOYANDNGVDGEMTYDDATKFTT 381
|||||
OY 361 VTEKPEVIDASELTPAVTYTKLVINGKTKLGETTTTAKVDAETAEAKAFKQYANDNGVDGYW 420
|||||
Db 382 VTEKPEVIDASELTPAVTYTKLVINGKTKLGETTTTAKVDAETAEAKAFKQYANDNGVDGYW 441
|||||
OY 421 TYDDATKFTTFTTEM 434
|||||
Db 442 TYDDATKFTTFTTEM 455
|||||

RESULT 2

Q51912 PRELIMINARY; PRT; 719 AA.
ID Q51912
AC Q51912;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PROTEIN L PRECURSOR.
OS Peptostreptococcus magnus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Finegoldia.
OX NCBI_TaxID=1260;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=312;
RX MEDLINE=90215984; PubMed=2108927;
RA Kastern W., Holst E., Nielsen E., Sjoerling U., Bjorck L.;
RT "Protein L, a bacterial immunoglobulin-binding protein and possible
RT virulence determinant";
RL Infect. Immun. 58:1217-1222(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=312;
RX MEDLINE=92316971; PubMed=1618782;
RA Bjoerck L., Sjoerling U., Kastern W.;
RT "Structure of peptostreptococcal protein L and identification of
RT repeated immunoglobulin light chain-binding domain";
RL J. Biol. Chem. 267:12820-12825(1992).
DR EMBL: M86697; AAA25612.1; -;
DR InterPro: IPR003147; B1.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF02246; B1; 5.
DR Pfam: PF00746; Gram_pos_anchor; 1.
KW Signal.
FT SIGNAL.
FT CHAIN 1 18 POTENTIAL.
FT CHAIN 19 719 PROTEIN L.
SQ SEQUENCE 719 AA; 78983 MW; 963A8D76D5E34DD2 CRC64;

Query Match 71.0%; Score 1587.5; DB 2; Length 719;
Best Local Similarity 76.3%; Pred. No. 3e-69;
Matches 334; Conservative 13; Mismatches 56; Indels 35; Gaps 6;

OY 2 VENKEETPTPTDSEEEVTIKANLIFANGSTQTAFFKGTFAEAYADTLKKDNG 61
|||||
Db 95 VENKEETPTPTDSEEEVTIKANLIFANGSTQTAFFKGTFAEAYADTLKKDNG 154
|||||
OY 62 EYTVADVADKGYTLNIRFAGKEKTPPEPKEEVTIKANLIYADGKTQTAFFKGTFAEATAEA 121
|||||

Db 155 EYTVADVADKGYTLNIRFAGKEKTPPEPKEEVTIKANLIYADGKTQTAFFKGTFAEATAEA 214
|||||
OY 122 YRYADALKKONGEYTVDVADKGYTLNIRFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 181
|||||
Db 215 YRYADALKKONGEYTVDVADKGYTLNIRFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 274
|||||
OY 182 FKGTTEETAEAYRYADLLAKENGKGYTVDVADKGYTLNIRFAGKEKTPPEPKEEVTIKAN 241
|||||
Db 275 FKGTTEETAEAYRYADLLAKENGKGYTVDVADKGYTLNIRFAGKEKTPPEPKEEVTIKAN 334
|||||
OY 242 LIYADGKTQTAFFKGTFAEAYRYADLLAKENGKGYTVADLEGGYTNIRFAGKKYD 301
|||||
Db 335 LIYADGKTQTAFFKGTFAEAYRYADLLAKENGKGYTVADLEGGYTNIRFAGKKYD 394
|||||
OY 302 KPEEPMGYTK - LILNGKTKLGETTTEAVDAATAEKYVFKOYANDNGVD - GEMTYDDATK 357
|||||
Db 395 KPEEKEQVTKENIYFEDGTGTATFAGTFAEATAEA - RYADLLSKERHGYTADLEGG 452
|||||
OY 358 TPTVT - - - - - EKPEVIDASELTPAVTYTKLVINGKTKLGETTTTAKVDAETAE 404
|||||
Db 453 GYTNIRFAGKEPPEPKEVQD - - - - - GYASVEEAEAAAKEALKND 496
|||||
OY 405 KAFKQYANDNGVDGEMTY 422
|||||
Db 497 DVNKSYYTIROGADGRYY 514
|||||

RESULT 3

Q51918 PRELIMINARY; PRT; 992 AA.
ID Q51918
AC Q51918;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PROTEIN L PRECURSOR.
OS Peptostreptococcus magnus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Finegoldia.
OX NCBI_TaxID=1260;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3316;
RX MEDLINE=95078460; PubMed=7987012;
RA Murphy J.P., Troewen A.R., Duggleby C.J.;
RT "Nucleotide sequence of the gene for peptostreptococcal protein L";
RL DNA Seq. 4:259-265(1994).
DR EMBL: L04466; AAA67503.1; -;
DR HSP; Q51911; IGAB.
DR InterPro: IPR003147; B1.
DR InterPro: IPR002988; GA.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF02246; B1; 4.
DR Pfam: PF01468; GA; 4.
DR Pfam: PF00746; Gram_pos_anchor; 1.
KW Signal.
FT SIGNAL.
FT CHAIN 1 24 POTENTIAL.
FT CHAIN 25 992 PROTEIN L.
SQ SEQUENCE 992 AA; 108700 MW; 9CFF5771578A5DCE CRC64;

Query Match 56.5%; Score 1263.5; DB 2; Length 992;
Best Local Similarity 58.8%; Pred. No. 1.8e-53;
Matches 281; Conservative 36; Mismatches 90; Indels 71; Gaps 12;

OY 2 VENKEETPTPTDSEEEVTIKANLIFANGSTQTAFFKGTFAEAYADTLKKDNG 61
|||||
Db 241 MERKUSEKETPE - PEEVEVTIKANLIFADGSTQNAFFKGTFAKAVSDAYADALKKDNG 298
|||||
OY 62 EYTVADVADKGYTLNIRFAGKEKTPPEPKEEVTIKANLIYADGKTQTAFFKGTFAEATAEA 121
|||||
Db 299 EYTVADVADKGYTLNIRFAGKEKTPPEPKEEVTIKANLIYADGKTQTAFFKGTFAEATAEA 358
|||||
OY 122 YRYADALKKONGEYTVDVADKGYTLNIRFAGKE - - KTPPEPKEEVTIKANLIYADGKTOT 179
|||||

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OW protein - protein search, using sw model

Run on: October 29, 2002, 09:25:41 ; Search time 20.636 Seconds
(without alignments)
2556.865 Million cell updates/sec

Title: US-08-325-278B-1
Perfect score: 1565
Sequence: 1 AVENKEETPEPTDSEEV.....GGYTINIRFAGKVKDEKPEE 305

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP. archaea:*
2: SP. bacteria:*
3: SP. fungi:*
4: SP. human:*
5: SP. invertebrate:*
6: SP. mammal:*
7: SP. mhc:*
8: SP. organelle:*
9: SP. phage:*
10: SP. plant:*
11: SP. rodent:*
12: SP. virus:*
13: SP. vertebrate:*
14: SP. unclassified:*
15: SP. virus:*
16: SP. bacteriophage:*
17: SP. archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1565	100.0	455	2	053291
2	1561	99.7	719	2	051912
3	1226.5	78.4	992	2	051918
4	167	10.7	664	2	053975
5	166	10.6	479	10	091F88
6	166	10.6	669	2	093EM8
7	151	9.6	448	10	09SKP0
8	149	9.5	448	10	09SKP0
9	148.5	9.5	2276	2	093TW6
10	141.5	9.0	1110	13	091255
11	141	9.0	1849	2	09S4K2
12	137.5	8.8	992	16	09WZ46
13	137.5	8.8	5458	5	09U459
14	136	8.7	506	2	053837
15	134.5	8.6	3198	5	09U8G8
16	132.5	8.5	737	16	09JYK4

17	132.5	8.5	17352	5	095YM2	095YM2 procambarius
18	131.5	8.4	891	2	004111	004111 enterococcu
19	131.5	8.4	1166	2	086489	086489 staphylococ
20	130.5	8.3	564	2	069188	069188 streptococ
21	130.5	8.3	581	2	033742	033742 streptococ
22	130.5	8.3	663	2	030874	030874 streptococ
23	130.5	8.3	701	2	09KK48	09KK48 streptococ
24	130.5	8.3	1141	16	099W46	099W46 staphylococ
25	130.5	8.3	1733	2	0932F7	0932F7 staphylococ
26	129.5	8.3	1171	2	09KMX6	09KMX6 staphylococ
27	129	8.2	938	16	0927R4	0927R4 listeria in
28	128.5	8.2	1119	13	P87344	P87344 theragra ch
29	128.5	8.2	701	2	09ROT5	09ROT5 streptococ
30	127.5	8.1	728	2	094782	094782 clostridium
31	127.5	8.1	717	2	09EY85	09EY85 enterococcu
32	127.5	8.1	843	2	047802	047802 enterococcu
33	127.5	8.1	913	13	013099	013099 xenopus lae
34	127	8.1	1035	5	021380	021380 caenorhabd
35	126.5	8.1	1035	5	021380	021380 streptococ
36	126.5	8.1	2045	16	09AOK5	09AOK5 bacillus sp
37	126	8.1	955	2	045574	045574 glycoline max
38	125.5	8.0	673	10	02M9S5	02M9S5 caenorhabd
39	125.5	8.0	1510	5	025920	025920 plasmodium
40	125	8.0	463	10	039871	039871 mycoplasma
41	125	8.0	2083	5	09N435	09N435 mycoplasma
42	124.5	8.0	1302	2	049547	049547 mycoplasma
43	124	7.9	495	16	099MG8	099MG8 staphylococ
44	124	7.9	506	2	03834	03834 salmonella
45	124	7.9	1489	10	096449	096449 phytophthor

ALIGNMENTS

RESULT 1

053291 PRELIMINARY: PRT: 455 AA.

AC 053291;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PROTEIN IG (FRAGMENT).
OS Streptococcus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1306;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE-93094283; PubMed-1460053;
RA Kihlberg B.M., Sjoberg U., Kasten W., Bjorck L.,
RT "Protein Ig: a hybrid molecule with unique immunoglobulin binding properties."
RT J. Biol. Chem. 267:25583-25588 (1992).
DR EMBL: P06654; IPGX.
DR HSP: P06654; AA03280.1; .
DR InterPro: IPR003147; BL.
DR InterPro: IPR00724; Igc_bind_B.
DR Pfam: PF02246; BL: 4.
DR Pfam: PF01378; Igc_binding_B; 2.
FT NON_TER 455
SQ SEQUENCE 455 AA; 49926 MW; 381FC235B8C8307B CRC64;

QY	1	AVENKEETPEPTDSEEVITKANLIFANGSTOAEKGTPEKTSAYAYADLKDN	60
DB	22	AVENKEETPEPTDSEEVITKANLIFANGSTOAEKGTPEKTSAYAYADLKDN	81
QY	61	GEYTDVADKGYTLNIRKFAKKEKPEPEKTEVTIANLIVADGKTQTAKEFGTPEATAE	120
DB	82	GEYTDVADKGYTLNIRKFAKKEKPEPEKTEVTIANLIVADGKTQTAKEFGTPEATAE	141

```

OY 121 AYRYADALKKNGEYTVADKGYTLNIFAGKEKTPPEPKKEEVTIKANLIYADKGTCTA 180
DB 142 AYRYADALKKNGEYTVADKGYTLNIFAGKEKTPPEPKKEEVTIKANLIYADKGTCTA 201
OY 181 EFKGTFEEATARAAYRADLLAKENGKYYVADKGYTLNIFAGKEKTPPEPKKEEVTIKA 240
DB 202 EFKGTFEEATARAAYRADLLAKENGKYYVADKGYTLNIFAGKEKTPPEPKKEEVTIKA 261
OY 241 NLIYADGKTQTAEPFGTAFAATARAAYRADLLAKENGKYYTADLEGGYTINIRFAGKRYD 300
DB 262 NLIYADGKTQTAEPFGTAFAATARAAYRADLLAKENGKYYTADLEGGYTINIRFAGKRYD 321
OY 301 EKPEE 305
DB 322 EKPEE 326

RESULT 2
OY 051912 PRELIMINARY; PRT; 719 AA.
AC 051912:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PROTEIN L. PRECURSOR.
OS Peptostreptococcus magnus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Finegoldia.
OX NCBI_TaxID=1260;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=312;
RX MEDLINE=90215984; PubMed=2108927;
RA Kaster W., Holst E., Nielsen E., Stobring U., Bjorck L.;
RT "Structure of peptostreptococcal protein L and identification of
  repeated immunoglobulin light chain-binding domain."
RL J. Biol. Chem. 267:12820-12825(1992).
DR EMBL: M86697; AAA25612.1;
DR InterPro: IPR003147; B1.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF02246; B1; 5.
DR Signal.
KW SIGNAL.
FT CHAIN 1 18 POTENTIAL.
FT CHAIN 19 719 PROTEIN L.
SQ SEQUENCE 719 AA; 78983 MM; 963A8D76D5E34DD2 CRC64;

Query Match 99.7%; Score 1561; DB 2; Length 719;
Best Local Similarity 100.0%; Pred. No. 2.5e-78;
Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

DB 275 FKGTFEEATARAAYRADLLAKENGKYYVADKGYTLNIFAGKEKTPPEPKKEEVTIKAN 334
OY 242 LIYADGKTQTAEPFGTAFAATARAAYRADLLAKENGKYYTADLEGGYTINIRFAGKRYD 301
DB 335 LIYADGKTQTAEPFGTAFAATARAAYRADLLAKENGKYYTADLEGGYTINIRFAGKRYD 394
OY 302 KPEE 305
DB 395 KPEE 398

RESULT 3
OY 051918 PRELIMINARY; PRT; 992 AA.
AC 051918:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PROTEIN L. PRECURSOR.
OS Peptostreptococcus magnus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Finegoldia.
OX NCBI_TaxID=1260;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=3316;
RX MEDLINE=95078460; PubMed=7987012;
RA Murphy J.P., Trowern A.R., Dugdaley C.J.;
RT "Nucleotide sequence of the gene for peptostreptococcal protein L."
RL DNA Seq. 4:259-265(1994).
DR EMBL: I04466; AAA67503.1;
DR HSSP: Q05191; IGAB.
DR InterPro: IPR003147; B1.
DR InterPro: IPR002988; GA.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF02246; B1; 4.
DR Pfam: PF01468; GA; 4.
DR Signal.
KW SIGNAL.
FT CHAIN 1 24 POTENTIAL.
FT CHAIN 25 992 PROTEIN L.
SQ SEQUENCE 992 AA; 108700 MM; 9CFF5771578A5DCE CRC64;

Query Match 78.4%; Score 1226.5; DB 2; Length 992;
Best Local Similarity 79.9%; Pred. No. 8.5e-60;
Matches 246; Conservative 22; Mismatches 31; Indels 9; Gaps 4.

```